SKEW-NORMALITY IN STOCHASTIC FRONTIER ANALYSIS

J. Armando Domínguez-Molina, Graciela González-Farías and Rogelio Ramos-Quiroga

Comunicación Técnica No I-03-18/06-10-2003
(PE/CIMAT)

Skew-Normality in Stochastic Frontier Analysis

J. Armando Domínguez-Molina
Universidad de Guanajuato, México
Callejón de Jalisco s/n
36240 Guanajuato, MÉXICO
jadguez@cimat.mx
Graciela González-Farías
Centro de Investigación en Matemáticas A. C.
Callejón de Jalisco s/n
36240 Guanajuato, MÉXICO
farias@cimat.mx
and
Rogelio Ramos-Quiroga
Centro de Investigación en Matemáticas A. C.
Callejón de Jalisco s/n
36240 Guanajuato, MÉXICO
rramosq@cimat.mx

Abstract

Skewness is an intrinsic characteristic in Stochastic Frontier Analysis (SFA), where it is used as a measure of technical inefficiency. We discuss the use of skew normality in SFA. We consider stochastic frontier analysis in the common setting Normal + Truncated Normal with uncorrelated errors, as well as the case with correlated errors. In this last case we show the connection between the SFA model and the Closed Skew Normal as discussed in González-Farías, *et al* (2003). We end with the proposal of a model for stochastic frontier analysis with elliptical errors.

1 Introduction

A problem of interest to econometricians is the specification and estimation of a frontier production function. The original formulation of the stochastic frontier model is due to

Aigner, Lovell and Schmidt (1977) and Meeusen and van den Broeck (1977):

$$y = f(\mathbf{x}; \boldsymbol{\beta}) + e$$

where the error term $e = v - u$, is composed by a symmetric component, v, representing measurement error, and by the nonnegative technical inefficiency component u. The stochastic frontier $f(\mathbf{x}; \boldsymbol{\beta}) + v$ allows for firms with same inputs, x, to have different frontiers due to unobservable shocks; so, the model $y = f(\mathbf{x}; \boldsymbol{\beta}) + v - u$, models indeed the inefficiency of a company to attain its production frontier. See Parsons (2002) for a treatment of Stochastic Frontier Analysis in marketing science.

Some issues of interest in stochastic frontier analysis are the estimation of the common frontier $f(\mathbf{x}; \boldsymbol{\beta})$, the estimation of the technical efficiency (typically related to $\tau_u^2 = var(u)$), and the estimation of the measurement error $\sigma^2 = var(v)$. Before we discuss these objectives, we take a closer look at the structure of the disturbance term e, and its relation to Azzalini's proposal (1985) of a skewed distribution. The distributional properties of y will be inherited from those of e by a translation of the location parameter via $f(\mathbf{x}; \boldsymbol{\beta})$.

Aigner, *et al* (1977) discuss the model where u has a positive half normal distribution and is independent of v which is assumed normally distributed. Stevenson (1980) generalizes this model by considering u to have a normal distribution with mean μ and variance σ^2 truncated below at zero. That is, when $\mu = 0$ we have the half normal distribution. From Aigner, *et al* (1977) we have that the density of $e = v - u$, with u and v independent and $v \sim N(0, \sigma^2)$ and $u \sim N^0(0, \tau^2)$, (that is, the distribution of u is positive half normal) is given by:

$$g(e) = 2\frac{1}{\alpha}\phi\left(\frac{e}{\alpha}\right)\Phi\left(-\frac{\lambda}{\alpha}e\right), \tag{1}$$

where $\alpha = \sqrt{\tau^2 + \sigma^2}$, $\lambda = \tau/\sigma$, $\phi(\cdot)$ and $\Phi(\cdot)$ denote the density function and the distribution function of a standard normal random variable, respectively.
Comparing this function with the skew normal density introduced by Azzalini (1985):

$$g(x) = 2\phi(x)\Phi(\delta x), \ \delta \in \Re. \tag{2}$$

we see that the density of the disturbance e is just the density of a scaled skew normal distribution; that is $e = \alpha x$ where x is a skew normal with $\delta = -\lambda$.

Densities related to Azzalini's can be traced back to the work of Birnbaum (1950). For a historical note see Remark 2.2 in Arnold and Beaver (2002a). Azzalini's work, however, was the first to fully study its properties and to give it its name (Skew Normal Distribution). Azzalini's work of 1985 was followed by several developments related to the skew normal, we can mention Henze (1986) which shows how to obtain it as the sum of two random variables, one of them sign free, and the other a positive one. Azzalini (1985) also generalizes the skew normal density as:

$$g(x) = \Phi^{-1}\left(\nu/\sqrt{1+\delta^2}\right)\phi(x)\Phi(\delta x + \nu). \tag{3}$$

Arnold and Beaver (2000, 2002a, 2002b) describe several procedures that lead to skew densities, all of them cover the skew normal as a special case. One of these procedures considers the distribution of random variables of the form:

$$y = u + \delta v(c)$$

where u and v can have arbitrary distributions and $v(c) = vI(v;c)$, where $I(v;c) = 1$ if $v \geq c$ and zero otherwise; they give expressions for the cases u and v arbitrary, u and v symmetric, u and v non independent.

The usual setting in Stochastic Frontier Analysis is when we consider models for optimal production functions, thus the construct for the disturbance is $e = v - u$; if the setting under consideration involves minimal cost frontiers, then the usual device is to switch the sign of u: $e = v + u$. One direct generalization (of no use for the moment, but will be useful in the next section) is to consider

$$e = v + \delta u,$$

where δ is fixed, u is a random variable truncated below at a positive constant c. The parameter δ would indicate the direction of asymmetry. The corresponding density is given by:

$$g(e) = \int_{-\infty}^{\infty} f(e - \delta u)\, h(u)\, du,$$

if $v \sim N(\mu, \sigma^2)$ and $u \sim N^c(\nu, \tau^2)$ (the notation $N^c(\nu, \tau^2)$ indicates that u has a $N(\nu, \tau^2)$ distribution truncated below at c, $c \in \Re$.) then the density is:

$$\begin{aligned} g(e) &= \frac{\Phi^{-1}\left(\frac{\nu - c}{\tau}\right)}{\sqrt{\sigma^2 + \delta^2\tau^2}} \phi\left(\frac{e - \mu - \delta\nu}{\sqrt{\sigma^2 + \delta^2\tau^2}}\right) \\ &\times \Phi\left[\frac{\delta\tau(e - \mu - \delta\nu)}{\sigma\sqrt{\sigma^2 + \delta^2\tau^2}} + \frac{(\nu - c)\sqrt{\delta^2\tau^2 + \sigma^2}}{\sigma\tau}\right]. \end{aligned} \tag{4}$$

where $\alpha = \sqrt{\tau^2 + \sigma^2}$ (see Example in Appendix).

If $e^\star = u - v$, $u \sim N(0, \sigma^2)$, $v \sim N^0(\nu, \tau^2)$, u independent of v, then using (4) with $\mu = c = 0$, and $\delta = -1$:

$$g(e^\star) = \frac{\Phi^{-1}\left(\frac{\nu}{\tau}\right)}{\alpha} \phi\left(\frac{e + \nu}{\alpha}\right) \Phi\left[\frac{-\tau(e + \nu)}{\sigma\alpha} + \frac{\nu\alpha}{\sigma\tau}\right].$$

which is the same as in Kumbhakar and Lovell (2000, eq. 3.2.46) after simplification, i.e.,

$$g(e^\star) = \frac{1}{\alpha}\phi\left(\frac{e + \nu}{\alpha}\right) \Phi\left(\frac{\nu}{\alpha\lambda} - \frac{e\lambda}{\alpha}\right) / \Phi\left(\frac{\nu}{\sigma}\right).$$

where $\lambda = \tau/\sigma$.

Also, observe that if we define the random variable $w = \frac{e - \mu - \delta\nu}{\sqrt{\sigma^2 + \delta^2\tau^2}}$ and $\delta^\sharp = \delta\tau/\sigma$, $\nu^\sharp = \frac{(\nu - c)\sqrt{\delta^2\tau^2 + \sigma^2}}{\sigma\tau}$ then the density of w is the same as (3).

The error structure, $e = v + \delta u$, has been examined under several distributional assumptions for v and u. See Arnold and Beaver (2002a, 2002b), Kumbhakar and Lovell (2000).

2 Estimation

Let us assume a data structure of cross sectional type, composed of independent observations on n firms: We have their output production levels and the corresponding values of exogenous variables collected at a fixed period of time. We will assume a model of the form

$$y_i = f(\mathbf{x}_i; \boldsymbol{\beta}) + e_i \quad \text{where} \quad e_i = v_i + \delta_i u_i,\ \delta_i \in \Re, \quad i = 1, \cdots, n.$$

In the next subsection we will state a series of distributional assumptions on the error vector $\mathbf{e} = (e_1, \cdots, e_n)'$. However, the estimation problems are as before: We are interested in the estimation of the production technology parameter $\boldsymbol{\beta}$ in $f(\mathbf{x}; \boldsymbol{\beta})$, and also in the prediction of the technical efficiency of each firm, this last problem implies that we need to separate the statistical noise from the technical inefficiency.

2.1 Model Assumptions

A standard set of assumptions in Stochastic Frontier Analysis would include: Measurement error, v_i, are independent random shocks $N(0, \sigma^2)$, u_i are i.i.d. $N^c(\nu; \tau^2)$ *i.e.*, $\nu = c = 0$ is the nonnegative half normal distribution; also an independence condition such as u_i and v_j independent for all i and j; inputs, $\mathbf{x}_i$ are known non stochastic variables.

In this work we propose a general model that contains submodels of the stochastic frontier model with Normal Errors + Truncated Normal, this is:

$$\mathbf{y} = \mathbf{f}(X; \boldsymbol{\beta}) + \mathbf{v} + D\mathbf{u}, \tag{5}$$

where $\mathbf{v} = (v_1, ..., v_n)' \sim N_n(\mathbf{0}, \Sigma)$, $\mathbf{u} = (u_1, ..., u_m)' \sim N_m^{\mathbf{c}}(\boldsymbol{\nu}, \Lambda)$, $m \geq n$, see Section 4.1 for the definition of $N_m^{\mathbf{c}}(\boldsymbol{\nu}, \Lambda)$. $D(m \times n)$ is a full row rank matrix. $\mathbf{v}$ independent of $\mathbf{u}$, $\mathbf{f}(\mathbf{x}; \boldsymbol{\beta}) = (f(\mathbf{x}_1; \boldsymbol{\beta}), ..., f(\mathbf{x}_n; \boldsymbol{\beta}))'$, $X = (\mathbf{x}_1, \cdots, \mathbf{x}_n)'$ a know matrix of covariables and $\boldsymbol{\beta}$ is unknown.

The matrix D gives flexibility to the model, if we leave it unspecified, we can estimate it and use this estimate to help validate the model assumptions, on the other hand, we can set it $D = I_n$ or $D = -I_n$ for efficiencies or inefficiencies, respectively. Finally, we can induce particular correlation structures in the errors by letting non-zero off-diagonal parameters.

From Theorem 1 we have that the density of $\mathbf{e} = \mathbf{v} + D\mathbf{u}$ is:

$$\begin{aligned} g(\mathbf{e}) = {} & \Phi_m^{-1}(\mathbf{0}; \mathbf{c} - \boldsymbol{\nu}, \Lambda)\, \phi_n(\mathbf{e}; D\boldsymbol{\nu}, \Sigma + D\Lambda D') \\ & \times \Phi_m \Big[\Lambda D' (\Sigma + D\Lambda D')^{-1} (\mathbf{e} - D\boldsymbol{\nu}); \mathbf{c} - \boldsymbol{\nu}, \\ & (D'\Sigma^{-1} D + \Lambda^{-1})^{-1}\Big], \end{aligned} \tag{6}$$

where $\phi_p(\cdot\,;\boldsymbol{\mu},\Sigma)$ and $\Phi_p(\cdot\,;\boldsymbol{\mu},\Sigma)$ denote the p.d.f. and the c.d.f. of a p-dimensional normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix Σ, respectively.
Thus $\mathbf{e}$ has a closed skew normal distribution, i.e.,

$$\mathbf{e} \sim CSN_{n,m}\left(\boldsymbol{\mu}^{\star}, \Sigma^{\star}, D^{\star}, \boldsymbol{\nu}^{\star}, \Delta^{\star}\right)$$

where $\boldsymbol{\mu}^{\star} = D\boldsymbol{\nu},\ \Sigma^{\star} = \Sigma + D\Lambda D',\ D^{\star} = \Lambda D'\left(\Sigma + D\Lambda D'\right)^{-1},\ \boldsymbol{\nu}^{\star} = \mathbf{c} - \boldsymbol{\nu}$. See equation (11) and González-Farías, *et al* (2003a,b) and Domínguez-Molina, *et al* (2003).

This model includes the following cases as submodels

Model I: homoscedastic and uncorrelated errors.
If we set in the model (5) $D = \delta I_n, \Sigma = \sigma^2 I_n, \Lambda = \tau^2 I_n$, where I_n is the $n \times n$ identity matrix, we would have the case of homoscedastic and uncorrelated observations: $var(y_i) = constant$, $i = 1, 2, \cdots, n$ and $cov(y_i, y_j) = cov(e_i, e_j) = 0$, $i \neq j$, $i, j = 1, 2, \cdots, n$.

Model II: heteroscedastic and uncorrelated errors.
If the matrices D, Σ and Λ are diagonal and if any of them is of the form

$$D = diag(\delta_1, ..., \delta_n),\ \Sigma = diag\left(\sigma_1^2, ..., \sigma_n^2\right),\ \Lambda = diag\left(\tau_1^2, ..., \tau_n^2\right),$$

then we would have the case of heteroscedastic but uncorrelated observations: $var(y_i) = k_i$, $i = 1, 2, \cdots, n$ and $k_i \neq k_j$ for some $i \neq j$, and $cov(y_i, y_j) = cov(e_i, e_j) = 0$, $i \neq j$, $i, j = 1, 2, \cdots, n$.

Model III: correlated errors.
If any of the matrices D, Σ or Λ are non-diagonal we would have the case of correlated errors $cov(y_i, y_j) = cov(e_i, e_j) \neq 0$ for some $i \neq j$. That is $cov(\mathbf{e}) = \Omega$, could be non-diagonal. If in this submodel we have:

$$\Sigma > 0,\ D = diag(\delta_1, ..., \delta_n)\ ,\ \Lambda = diag\left(\tau_1^2, ..., \tau_n^2\right),$$

then the errors are correlated and the marginal distribution of the errors would easily computable, because in this case $v_i \sim N(0, \sigma_i^2)$ and $u_i \sim N^c(\nu_i; \tau_i^2)$ and independent and we can use (4) in order to evaluate the marginal distributions of the errors.

For our general model (5) using Remark 1 of González-Farías, *et al* (2003a), we have that the marginal distribution of the errors is:

$$e_i \sim CSN_{1,m}\left(\mathbf{0}, \Sigma^{\star}_{\mathbf{a}_i}, D^{\star}_{\mathbf{a}_i}, \boldsymbol{\nu}, \Delta^{\star}_{\mathbf{a}_i}\right),$$

where:

$$\begin{aligned}
\Sigma^{\star}_{\mathbf{a}_i} &= \mathbf{a}_i'\Sigma^{\star}\mathbf{a}_i \\
D^{\star}_{\mathbf{a}_i} &= D^{\star}\Sigma^{\star}\mathbf{a}_i\Sigma^{\star -1}_{\mathbf{a}_i} \\
\Delta^{\star}_{\mathbf{a}_i} &= \Delta^{\star} + D^{\star}\Sigma^{\star}D^{\star\prime} - D^{\star}\Sigma^{\star}\mathbf{a}_i\mathbf{a}_i'\Sigma^{\star}D^{\star\prime}\Sigma^{\star -1}_{\mathbf{a}_i},
\end{aligned}$$

and $\mathbf{a}_i$ is the *i-th* unit vector in $\Re^n$.

2.2 Likelihood

Given a data structure of a vector of observations $\mathbf{y} = (y_1, ..., y_n)'$ and the corresponding set of inputs $X = (\mathbf{x}_1, \cdots, \mathbf{x}_n)'$, then under the assumptions of model (5) we can write the likelihood function and base our inferences on it.

From (6) and (5) we have that the likelihood function of the parameters $\boldsymbol{\beta}, \Sigma, D, \boldsymbol{\nu}, \mathbf{c}, \Lambda$ is:

$$
\begin{aligned}
& L\left(\boldsymbol{\beta}, \Sigma, D, \boldsymbol{\nu}, \mathbf{c}, \Lambda\right) \\
& = \Phi_m^{-1}\left(\mathbf{0}; \mathbf{c} - \boldsymbol{\nu}, \Lambda\right) \phi_n\left(\mathbf{y} - \mathbf{f}\left(\mathbf{x}; \boldsymbol{\beta}\right); D\boldsymbol{\nu}, \Sigma + D\Lambda D'\right) \\
& \times \Phi_m \left\{ \Lambda D'\left(\Sigma + D\Lambda D'\right)^{-1}\left[\mathbf{y} - \mathbf{f}\left(\mathbf{x}; \boldsymbol{\beta}\right) - D\boldsymbol{\nu}\right]; \right. \\
& \left. \mathbf{c} - \boldsymbol{\nu}, \left(D'\Sigma^{-1}D + \Lambda^{-1}\right)^{-1} \right\}.
\end{aligned}
$$

For the non-correlated case $\Sigma = diag\left(\sigma_1^2, ..., \sigma_n^2\right),\ D = diag\left(\delta_1, ..., \delta_n\right),\ \Lambda = diag\left(\tau_1^2, ..., \tau_n^2\right)$ the likelihood reduces to:

$$
\begin{aligned}
& L\left(\boldsymbol{\beta}, \Sigma, D, \boldsymbol{\nu}, \mathbf{c}, \Lambda\right) \\
& = \prod_{i=1}^{n} \left\{ \Phi^{-1}\left(\frac{v_i - c_i}{\tau_i}\right) \phi\left(y_i - f\left(\mathbf{x}_i; \boldsymbol{\beta}\right); \delta_i\nu_i, \sigma_i^2 + \delta_i^2\tau_i^2\right) \right. \\
& \times \Phi\left[\tau_i^2\delta_i\left(\sigma_i^2 + \delta_i^2\tau_i^2\right)^{-1}\left(y_i - f\left(\mathbf{x}_i; \boldsymbol{\beta}\right) - \delta_i\nu_i\right); \right. \\
& \left.\left. c_i - \nu_i, \tau_i^2 - \tau_i^4\delta_i^2\left(\sigma_i^2 + \delta_i^2\tau_i^2\right)^{-1}\right]\right\} \\
& = \prod_{i=1}^{n} \left\{ \frac{\Phi^{-1}\left(\frac{v-c}{\tau_i}\right)}{\sqrt{\sigma_i^2 + \delta_i^2\tau_i^2}} \phi\left(\frac{y_i - f\left(\mathbf{x}_i; \boldsymbol{\beta}\right) - \delta_i\nu_i}{\sqrt{\sigma_i^2 + \delta_i^2\tau_i^2}}\right) \right. \\
& \left. \times \Phi\left(\frac{\tau_i\delta_i\left(y_i - f\left(\mathbf{x}_i; \boldsymbol{\beta}\right) - \delta_i\nu_i\right)}{\sigma_i\sqrt{\sigma_i^2 + \delta_i^2\tau_i^2}}; \left(c_i - \nu_i\right)\frac{\sqrt{\sigma_i^2 + \delta_i^2\tau_i^2}}{\tau_i\sigma_i}\right)\right\}.
\end{aligned}
$$

2.3 Estimation of inefficiencies/efficiencies

In the univariate model of SFA the error is of the form $e = v + \delta u$. The main interest in SFA is to estimate inefficiencies δu $(\delta < 0)$ or efficiencies δu $(\delta > 0)$ for each x.
After the estimation of $\boldsymbol{\beta}$ by maximum likelihood we can evaluate the residuals:

$$\hat{e}_i = y_i - f\left(\mathbf{x}_i; \widehat{\boldsymbol{\beta}}\right).$$

The most common proposals for calculating the inefficiencies are by means of the mean or the mode of the random variable $u|e$, that is:

$$\text{efficiency}\left(\boldsymbol{\beta}, \sigma, \delta, \nu, \tau, c, e; \mathbf{x}\right) = \delta E\left(u|e\right),\ \text{or}$$

$$\text{efficiency}\left(\boldsymbol{\beta},\sigma,\delta,\nu,\tau,c,e;\mathbf{x}\right)=\delta M\left(u|e\right),$$

where $M(\cdot)$ is the mode. Thus we estimate the efficiency of firm i by:

$$\text{efficiency}_i^*\left(\widehat{\boldsymbol{\beta}},\hat{\sigma},\hat{\delta},\hat{\nu},\hat{\tau},\hat{c},\hat{e}_i;\mathbf{x}_i\right).$$

For our model (5) we proceed in a similar way. Let us assume $\mathbf{v}\sim N_n\left(\mathbf{0},\Sigma\right)$ and $\mathbf{u}\sim N_m^c\left(\boldsymbol{\nu},\Lambda\right)$, $\mathbf{v}$ and $\mathbf{u}$ independent, then if

$$\mathbf{e}=\mathbf{v}+D\mathbf{u},$$

then the joint density of $\mathbf{u}$ and $\mathbf{e}$ is given by:

$$g\left(\mathbf{u},\mathbf{e}\right)=\Phi_m^{-1}\left(-\mathbf{c};-\boldsymbol{\nu},\Lambda\right)\phi_n\left(\mathbf{e}-D\mathbf{u};\mathbf{0},\Sigma\right)\phi_m\left(\mathbf{u};\boldsymbol{\nu},\Lambda\right).$$

which, after some manipulations and the help of eq. A.2.4f of Mardia (1972, p458), it can be reduced to,

$$\begin{aligned} g\left(\mathbf{u},\mathbf{e}\right)=\phi_m\Big[&\mathbf{u};\boldsymbol{\nu}+\left(D'\Sigma^{-1}D+\Lambda^{-1}\right)^{-1}D'\Sigma^{-1}\left(\mathbf{e}-D\boldsymbol{\nu}\right),\\ &\left(D'\Sigma^{-1}D+\Lambda^{-1}\right)^{-1}\Big]\phi_n\left(\mathbf{e};D\boldsymbol{\nu},\Sigma+D'\Lambda D\right). \end{aligned} \tag{7}$$

From (10) and (6) we have that:

$$\begin{aligned} g\left(\mathbf{u}|\mathbf{e}\right)&=g\left(\mathbf{u},\mathbf{e}\right)/g\left(\mathbf{e}\right)\\ &=K\phi_m\Big[\mathbf{u};\boldsymbol{\nu}+\left(D'\Sigma^{-1}D+\Lambda^{-1}\right)^{-1}D'\Sigma^{-1}\left(\mathbf{e}-D\boldsymbol{\nu}\right),\\ &\left(D'\Sigma^{-1}D+\Lambda^{-1}\right)^{-1}\Big], \end{aligned}$$

where:

$$\begin{aligned} K^{-1}=\Phi_m\Big[&\boldsymbol{\nu}+\left(D'\Sigma^{-1}D+\Lambda^{-1}\right)^{-1}D'\Sigma^{-1}\left(\mathbf{e}-D\boldsymbol{\nu}\right);\mathbf{c},\\ &\left(D'\Sigma^{-1}D+\Lambda^{-1}\right)^{-1}\Big], \end{aligned}$$

that is:

$$\begin{aligned} \mathbf{u}|\mathbf{e}\sim N_m^{\mathbf{c}}\Big[&\boldsymbol{\nu}+\left(D'\Sigma^{-1}D+\Lambda^{-1}\right)^{-1}D'\Sigma^{-1}\left(\mathbf{e}-D\boldsymbol{\nu}\right),\\ &\left(D'\Sigma^{-1}D+\Lambda^{-1}\right)^{-1}\Big]. \end{aligned}$$

Domínguez-Molina, *et al* (2003) illustrate the procedure to obtain the mean of a Closed Skew Normal distribution, which can be useful to obtain the mean of $\mathbf{u}|\mathbf{e}$ ($E(\mathbf{u}|\mathbf{e})$) then we can obtain an estimation of the inefficiencies/efficiencies in our model, substituting the corresponding estimates in:

$$\text{efficiency}\left(\boldsymbol{\beta},\Sigma,D,\boldsymbol{\nu},\Lambda,\mathbf{c},\mathbf{e};\mathbf{x}\right)=DE\left(\mathbf{u}|\mathbf{e}\right).$$

See also Kotz, *et al* (2000), section 45.10.

3 SFA with skew elliptical components

Definition. A p-vector $\mathbf{w}$ has a spherical distribution if and only if for its characteristic function $\psi(\mathbf{t})$ there exist a function $\varphi(\cdot)$ of scalar variable such that $\varphi(\mathbf{t}) = \psi(\mathbf{t}'\mathbf{t})$. The function $\psi(\cdot)$ is called the *characteristic generator.*
Definition. A $p \times 1$ random vector v is said to have an *elliptical distribution (*or *elliptically symmetric* distribution) with parameters $\boldsymbol{\mu} \in \Re^p$ and $\Sigma\,(p \times p)$ if:

$$\mathbf{v} = \boldsymbol{\mu} + A'\mathbf{w},$$

where $\mathbf{w}$ has spherical distribution with characteristic generator ψ. Where $A\,(k \times p)$, $A'A = \Sigma$ with rank$(\Sigma) = k$. We will write $\mathbf{v} \sim EC_p(\boldsymbol{\mu}, \Sigma; \psi)$.

A random vector $\mathbf{v} \sim EC_p(\boldsymbol{\mu}, \Sigma; \psi)$, in general, does not necessarily possess a density. It is however possible to show that the density of $\mathbf{v}$, if it exists, must be of the form $h(\mathbf{v}'\mathbf{v})$ for some nonnegative function $h(\cdot)$ of a scalar variable such that:

$$\int_0^\infty y^{\frac{p}{2}-1} h(y)\, dy < \infty.$$

In this case we will write $\mathbf{v} \sim EC_p(\mu, \Sigma; h)$. See Fang, *et al* (1990, p35).

Consider the stochastic frontier model:

$$\mathbf{y} = \mathbf{f}(X; \boldsymbol{\beta}) + \mathbf{v} + D\mathbf{u},$$

where $\mathbf{v} = (v_1, ..., v_n)' \sim EC_n(\mathbf{0}, \Sigma; h_1)$, $\mathbf{u} = (u_1, ..., u_m)' \sim EC_m^{\mathbf{c}}(\boldsymbol{\nu}, \Lambda; h_2)$, $m \geq n$, see (9), $\mathbf{v}$ independent of $\mathbf{u}$, D is a full row rank matrix, $\mathbf{f}(X; \boldsymbol{\beta}) = (f(\mathbf{x}_1; \boldsymbol{\beta}), ..., f(\mathbf{x}_n; \boldsymbol{\beta}))'$, and $X = (\mathbf{x}_1, ..., \mathbf{x}_n)'$, $\mathbf{x}_i$ is a known vector of $\Re^{d_1}$, d_1 a positive integer, $i = 1, 2, ..., n$. and the unknown. $\boldsymbol{\beta} \in \Re^{d_2}$, d_2 a positive integer.
The density function of $\mathbf{e}$ is of the form:

$$g(\mathbf{e}) = K \int_{c_1}^\infty \cdots \int_{c_n}^\infty h_1\left[(\mathbf{e} - \mathbf{Du})'(\mathbf{e} - \mathbf{Du})\right] h_2(\mathbf{u})\, d\mathbf{u},$$

where K is a normalizing constant. This expression, in general, is difficult to evaluate. If the moment generating functions of $\mathbf{u}$ and $\mathbf{v}$ are available, then the m.g.f. of $\mathbf{e}$ is of the form:

$$M_{\mathbf{e}}(\mathbf{t}) = M_{\mathbf{v}}(\mathbf{t})\, M_{\mathbf{u}}(D'\mathbf{t}), \tag{8}$$

This expression could be useful to get the density function of $\mathbf{e}$ or, alternatively, we could use it to obtain the moments of $\mathbf{e}$ and use a method of moments to estimate the parameters of the model.

Acknowledgements

This work was supported by the research project 39017-E of CONACYT- México and the first two authors have been partially supported the project 03-16-K118-027 (Anexo 01) of CONCYTEG.

4 Appendix

4.1 Linear combination of a normal random vector and a truncated normal random vector

Let $\mathbf{w}=(w_1,...,w_p)'$, we consider truncation of the type $\mathbf{w}\geq\mathbf{c}$, where $\mathbf{w}\geq\mathbf{c}$ means $w_j\geq c_j$, $j=1,...,p$, that is, values of w_j less than c_j are excluded. (see Kotz, *et al*, 2000, Section 45.10).
For arbitrary vector $\mathbf{c}\in\Re^p$ define the function:

$$I(\mathbf{x};\mathbf{c})=\begin{cases}\mathbf{x} \text{ if } \mathbf{x}\geq\mathbf{c},\\ \mathbf{0} \text{ if otherwise.}\end{cases}$$

Let $\mathbf{w}$ a random vector, we denote by $\mathbf{w}(\mathbf{c})$ a random vector with truncation of the type $\mathbf{w}\geq\mathbf{c}$, for $\mathbf{c}\in\Re^p$ then $\mathbf{w}(\mathbf{c})$ can be written as:

$$\mathbf{w}(\mathbf{c})=\mathbf{w}I(\mathbf{w};\mathbf{c}), \tag{9}$$

thus, if $\mathbf{w}$ has density f, the density function of $\mathbf{w}(\mathbf{c})$ is given by:

$$g(\mathbf{w})=\frac{f(\mathbf{w})}{\Pr(\mathbf{w}\geq\mathbf{c})}I(\mathbf{w};\mathbf{c}).$$

If $\mathbf{w}\sim EC_p(\boldsymbol{\mu},\Sigma;h)$ and $\mathbf{u}\overset{D}{=}\mathbf{w}(\mathbf{c})$, where $\overset{D}{=}$ means equality in distribution, then the density function of $\mathbf{u}$ is:

$$\begin{aligned}g(\mathbf{u})&=\frac{f(\mathbf{u};\boldsymbol{\mu},\Sigma)}{\Pr(\mathbf{u}\geq\mathbf{c})}I(\mathbf{u};\mathbf{c})\\&=\frac{f(\mathbf{u};\boldsymbol{\mu},\Sigma)}{\Pr(-\mathbf{u}\leq-\mathbf{c})}I(\mathbf{u};\mathbf{c})\\&=\frac{f(\mathbf{u};\boldsymbol{\mu},\Sigma)}{F(-\mathbf{c};-\boldsymbol{\mu},\Sigma)}I(\mathbf{u};\mathbf{c}),\end{aligned}$$

where f and F are the density and the distribution function of $\mathbf{w}$, respectively.
If $\mathbf{w}\sim\mathbf{EC}_p(\boldsymbol{\mu},\Sigma;h)$, $h(w)=e^{-w/2}$, that is $\mathbf{w}\sim N_p(\boldsymbol{\mu},\Sigma)$, and $\mathbf{u}\overset{D}{=}\mathbf{w}(\mathbf{c})$ thus:

$$\begin{aligned}g(\mathbf{u})&=[\Pr(\mathbf{x}\geq\mathbf{c})]^{-1}\phi_p(\mathbf{u};\boldsymbol{\mu},\Sigma)I(\mathbf{u};\mathbf{c})\\&=[\Pr(-\mathbf{x}\leq-\mathbf{c})]^{-1}\phi_p(\mathbf{u};\boldsymbol{\mu},\Sigma)I(\mathbf{u};\mathbf{c})\\&=\Phi_p^{-1}(-\mathbf{c};-\boldsymbol{\mu},\Sigma)\phi_p(\mathbf{u};\boldsymbol{\mu},\Sigma)I(\mathbf{u};\mathbf{c}).\end{aligned}$$

We will denote by $\mathbf{u}\sim EC_p^{\mathbf{c}}(\boldsymbol{\mu},\Sigma;h)$ if $\mathbf{u}\overset{D}{=}\mathbf{w}(\mathbf{c})$ and $\mathbf{w}\sim\mathbf{EC}_p(\boldsymbol{\mu},\Sigma;h)$, similarly $\mathbf{u}\sim N_p^{\mathbf{c}}(\boldsymbol{\mu},\Sigma)$ if $\mathbf{u}\overset{D}{=}\mathbf{w}(\mathbf{c})$ and $\mathbf{w}\sim N_p(\boldsymbol{\mu},\Sigma)$.

Theorem 1 *If* $\mathbf{v}\sim N_p(\boldsymbol{\mu},\Sigma)$ *and* $\mathbf{u}\sim N_q^c(\boldsymbol{\nu},\Lambda)$, *and* $\mathbf{v}$ *is independent of* $\mathbf{u}$ *and:*

$$\mathbf{e}=\mathbf{v}+D\mathbf{u},$$

where D is a full row rank matrix, then

$$\mathbf{e} \sim CSN_{p,q}\left(\boldsymbol{\mu}^{\dagger}, \Sigma^{\dagger}, D^{\dagger}, \boldsymbol{\nu}^{\dagger}, \Delta^{\dagger}\right)$$

where $\mu^{\dagger} = \boldsymbol{\mu} + D\boldsymbol{\nu}$, $\Sigma^{\dagger} = \Sigma + D\Lambda D'$, $D^{\dagger} = \Lambda D'\left(\Sigma + D\Lambda D'\right)^{-1}$, $\boldsymbol{\nu}^{\dagger} = \mathbf{c} - \boldsymbol{\nu}$, $\Delta^{\dagger} = \left(D'\Sigma^{-1}D + \Lambda^{-1}\right)^{-1}$.
That is, the density function of $\mathbf{e}$ *is:*

$$\begin{aligned} g(\mathbf{e}) = {} & \Phi_p^{-1}\left(\mathbf{0}; \mathbf{c} - \boldsymbol{\nu}, \Lambda\right)\phi_p\left(\mathbf{e}; \boldsymbol{\mu} + D\boldsymbol{\nu}, \Sigma + D\Lambda D'\right) \\ & \times \Phi_p\left[\Lambda D'\left(\Sigma + D\Lambda D'\right)^{-1}\left(\mathbf{e} - \boldsymbol{\mu} + D\boldsymbol{\nu}\right); \mathbf{c} - \boldsymbol{\nu}, \right. \\ & \left. \left(D'\Sigma^{-1}D + \Lambda^{-1}\right)^{-1}\right]. \end{aligned} \tag{10}$$

Before proving Theorem 1 we will define the closed skew normal (CSN) distribution. González-Farías, *et al* (2003b) defines a random vector, $\mathbf{y}$, to have the CSN distribution if its density function is given by:

$$g_{p,q}(\mathbf{y}) = C\phi_p\left(\mathbf{y}; \boldsymbol{\mu}, \Sigma\right)\Phi_q\left[D\left(\mathbf{y} - \boldsymbol{\mu}\right); \boldsymbol{\nu}, \Delta\right], \quad \mathbf{y} \in \Re^p, \tag{11}$$

where:

$$C^{-1} = \Phi_q\left(\mathbf{0}; \boldsymbol{\nu}, \Delta + D\Sigma D'\right), \tag{12}$$

where $\phi_p(\cdot\,; \boldsymbol{\mu}, \Sigma)$ and $\Phi_p(\cdot\,; \boldsymbol{\mu}, \Sigma)$ denote the p.d.f. and the c.d.f. of a p-dimensional normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix Σ, respectively. We denote this as:

$$\mathbf{y} \sim CSN_{p,q}\left(\boldsymbol{\mu}, \Sigma, D, \boldsymbol{\nu}, \Delta\right).$$

The moment generating function of $\mathbf{y}$ is:

$$M_{\mathbf{y}}(\mathbf{t}) = \frac{\Phi_q\left(D\Sigma\mathbf{t}; \boldsymbol{\nu}, \Delta + D\Sigma D'\right)}{\Phi_q\left(\mathbf{0}; \boldsymbol{\nu}, \Delta + D\Sigma D'\right)} e^{\mathbf{t}' \ \ + \frac{1}{2}\mathbf{t}'\Sigma\mathbf{t}}, \quad \mathbf{t} \in \Re^p. \tag{13}$$

Lemma 2 *If* $\mathbf{w} \sim N_q(\boldsymbol{\nu}, \Lambda)$ *then the moment generating function of* $\mathbf{u} = \mathbf{w}(\mathbf{c})$ *is given by:*

$$M_{\mathbf{u}}(\mathbf{t}) = \Phi_q^{-1}\left(\mathbf{0}; \mathbf{c} - \boldsymbol{\nu}, \Lambda\right) e^{\mathbf{t}' \ \ + \frac{1}{2}\mathbf{t}'\Lambda\mathbf{t}} \Phi_q\left(\Lambda\mathbf{t}; \mathbf{c} - \boldsymbol{\nu}, \Lambda\right).$$

Proof.

$$\begin{aligned} M_{\mathbf{u}}(\mathbf{t}) &= Ee^{\mathbf{t}'\mathbf{u}} \\ &= \Phi_q^{-1}\left(-\mathbf{c}; -\boldsymbol{\nu}, \Lambda\right) \\ &\times \int_{c_1}^{\infty} \cdots \int_{c_q}^{\infty} e^{\mathbf{t}'\mathbf{u}} \phi_q\left(\mathbf{u}; \boldsymbol{\nu}, \Lambda\right) dw_1 \cdots dw_q \end{aligned}$$

given that:

$$e^{\mathbf{t}'\mathbf{u}}\phi_p\left(\mathbf{u}; \boldsymbol{\nu}, \Lambda\right) = e^{t' \ \ + \frac{1}{2}\mathbf{t}'\Lambda\mathbf{t}}\phi_p\left(\mathbf{u}; \boldsymbol{\nu} + \Lambda\mathbf{t}, \Lambda\right),$$

we get:

$$\begin{aligned} M_{\mathbf{u}}(\mathbf{t}) &= \Phi_q^{-1}(-\mathbf{c}; -\boldsymbol{\nu}, \Lambda)\, e^{\mathbf{t}'\ \ +\frac{1}{2}\mathbf{t}'\Lambda\mathbf{t}} \\ &\quad \times \int_{c_1}^{\infty} \cdots \int_{c_q}^{\infty} \phi_q(\mathbf{u}; \boldsymbol{\nu} + \Lambda\mathbf{t}, \Lambda)\, dw_1 \cdots dw_q \\ &\quad = \Phi_q^{-1}(-\mathbf{c}; -\boldsymbol{\nu}, \Lambda)\, e^{\mathbf{t}'\ \ +\frac{1}{2}\mathbf{t}'\Lambda\mathbf{t}} \Phi_q(-\mathbf{c}; -\boldsymbol{\nu} - \Lambda\mathbf{t}, \Lambda). \end{aligned}$$

■

Proof of Theorem 1. From Lemma 2 and the fact that $M_{D\mathbf{u}}(\mathbf{t}) = M_{\mathbf{u}}(D'\mathbf{t})$ we obtain:

$$M_{D\mathbf{u}}(\mathbf{t}) = \Phi_q^{-1}(0; \mathbf{c} - \boldsymbol{\nu}, \Lambda)\, e^{\mathbf{t}'D\ \ +\frac{1}{2}\mathbf{t}'D\Lambda D'\mathbf{t}} \Phi_q(\Lambda D'\mathbf{t}; \mathbf{c} - \boldsymbol{\nu}, \Lambda)$$

and given that $\mathbf{w} \sim N_p(\boldsymbol{\mu}, \Sigma)$ the m.g.f. of $\mathbf{w}$ is given by:

$$M_{\mathbf{w}}(\mathbf{t}) = e^{\mathbf{t}'\ \ +\frac{1}{2}\mathbf{t}'\Sigma\mathbf{t}}.$$

By independence of $\mathbf{v}$ and $\mathbf{u}$ the m.g.f. of $\mathbf{v} + D\mathbf{u}$ is:

$$\begin{aligned} M_{\mathbf{e}}(\mathbf{t}) &= M_{\mathbf{v}+D\mathbf{u}}(\mathbf{t}) \\ &= M_{\mathbf{v}}(\mathbf{t})\, M_{\mathbf{u}}(D'\mathbf{t}) \\ &= \Phi_q^{-1}(\mathbf{0}; \mathbf{c} - \boldsymbol{\nu}, \Lambda)\, \Phi_q(\Lambda D'\mathbf{t}; \mathbf{c} - \boldsymbol{\nu}, \Lambda)\, e^{\mathbf{t}'(\ \ +D\ \ )+\frac{1}{2}\mathbf{t}'(\Sigma + D\Lambda D')\mathbf{t}} \\ &= \Phi_q^{-1}(\mathbf{0}; \mathbf{c} - \boldsymbol{\nu}, \Lambda) \\ &\quad \times \Phi_q\left[\Lambda D'(\Sigma + D\Lambda D')^{-1}(\Sigma + D\Lambda D')\,\mathbf{t}; \mathbf{c} - \boldsymbol{\nu}, \Lambda\right] \\ &\quad \times e^{\mathbf{t}'(\ \ +D\ \ )+\frac{1}{2}\mathbf{t}'(\Sigma + D\Lambda D')\mathbf{t}}. \end{aligned}$$

Comparing with the CSN m.g.f. (13) we get that the density function of $\mathbf{e}$ is:

$$g_{p,q}(\mathbf{e}) = C\phi_p\left(\mathbf{e}; \boldsymbol{\mu}^\dagger, \Sigma^\dagger\right) \Phi_q\left[D^\dagger\left(\mathbf{e} - \boldsymbol{\mu}^\dagger\right); \boldsymbol{\nu}^\dagger, \Delta^\dagger\right], \quad \mathbf{e} \in \Re^p,$$

where $\boldsymbol{\mu}^\dagger, \Sigma^\dagger, D^\dagger, \boldsymbol{\nu}^\dagger$ and $\Delta^\dagger$ are given in the statement of Theorem 1 This density could have been obtained by integrating (7) with respect to $\mathbf{u}$.

Example: (*The stochastic frontier error model*). If $v \sim N(\mu, \sigma^2)$ and $w \sim N^c(\nu, \tau^2)$, v and w are independent. Then the density function of:

$$e = v + \delta w,$$

is given by:

$$\begin{aligned} g(e) &= \Phi^{-1}\left(\frac{\nu - c}{\tau}\right) \phi\left(e; \mu + \delta\nu, \sigma^2 + \delta^2\tau^2\right) \\ &\quad \times \Phi\left[\frac{\delta\tau^2(e - \mu - \delta\nu)}{\sigma\tau\sqrt{\sigma^2 + \delta^2\tau^2}} + \frac{(\nu - c)\sqrt{\delta^2\tau^2 + \sigma^2}}{\sigma\tau}\right]. \end{aligned}$$

Proof. From Theorem 1 with $\Sigma = \sigma^2, D = \delta, \Lambda = \tau^2$ and the identities:

$$\mu^{\dagger} = \mu + \delta\nu, \Sigma^{\dagger} = \sigma^2 + \delta^2\tau^2,$$
$$D^{\dagger} = \Lambda D' \left(\Sigma + D\Lambda D'\right)^{-1} = \delta\tau^2 \left(\sigma^2 + \delta^2\tau^2\right)^{-1}, \nu^{\dagger} = c - \nu$$

and:

$$\left(D'\Sigma^{-1}D + \Lambda^{-1}\right)^{-1} = \left(\delta^2\sigma^{-2} + \tau^{-2}\right)^{-1} = \frac{\sigma^2\tau^2}{\delta^2\tau^2 + \sigma^2}$$

we obtain that the density function of e is:

$$g(e) = \Phi^{-1}\left(\frac{\nu - c}{\tau}\right)\phi\left(e; \mu + \delta\nu, \sigma^2 + \delta^2\tau^2\right)$$
$$\times \Phi\left[\delta\tau^2\left(\sigma^2 + \delta^2\tau^2\right)^{-1}(e - \mu - \delta\nu); c - \nu, \frac{\sigma^2\tau^2}{\delta^2\tau^2 + \sigma^2}\right],$$

simplifying the former expression we get the stated density.

References

[1] Aigner, D., Lovell, C.A.K. and Schmidt, P. (1977). Formulation of stochastic frontier production function model. J*ournal of Econometrics,* **6**, 21-37.

[2] Arnold, B.C. and Beaver, R.J. (2000). Some skewed multivariate distributions. *American Journal of Mathematical and Management Sciences,* ***20****, pp.27-38.*

[3] Arnold, B.C., Beaver, R.J. (2002a). Skew multivariate models related to hidden truncation and/or selective reporting. *Test. Sociedad de Estadística e Investigación Operativa.* Vol. **11,** No. 1, pp. 7-54.

[4] Arnold, B.C. and Beaver, R.J. (2002b). Alternative construction of skewed multivariate distributions. Technical Report 270, Department of Statistics, University of California, Riverside.

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

[6] Birnbaum, Z.W. (1950). Effect of linear truncation on a multinormal population. *The Annals of Mathematical Statistics*, **21**, 272-279.

[7] Domínguez-Molina, J. A., González-Farías, G. and Gupta, A.K. (2003) The Multivariate Closed Skew Normal Distribution. Department of Mathematics and Statistics, Bowling Green State University, Technical Report No. 03-12.

[8] Fang, K.-T., Kotz, S. and Ng, K.-W. (1990). *Symmetric multivariate and related distributions.* Monographs on Statistics and Applied Probability 55. Monographs on Statistics and Applied Probability 36. Chapman and Hall.

[9] González-Farías, G., Domínguez-Molina, J.A. and Gupta, A.K. (2003a). Additive properties of skew normal random vectors. To appear, *Journal of Statistical Planning and Inference.*

[10] González-Farías, G., Domínguez-Molina, J. A. and Gupta, A.K. (2003b) The Multivariate Closed Skew Normal Distribution. *(Genton M. editor) Edited volume on "Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality"*

[11] Henze, N. (1986). A probabilistic representation of the 'skew normal' distribution. *Scand. J. Statist.* 13, 271-275.

[12] Kotz, S, Balakrishnan, N. and Johnson, N.L. (2000). *Continuous Multivariate Distributions.* Volume 1: Models and Applications, 2nd Edition. Wiley Series in Probability and Statistics.

[13] Kumbhakar, S.C. and Lovell, C.A. (2000). *Stochastic frontier analysis.* Cambridge: Cambridge University Press.

[14] Mardia, K.V., Kent, J.T. and Bibby, J.M. (1979). *Multivariate Analysis.* Academic Press.

[15] Meeusen, W. and Van Den Broeck, J. (1977). Efficiency estimation from Cobb-Douglaas production functions with composed error. *International Economic Review,* Vol. 18, No. 2, pp. 435-444.

[16] Parsons, L.J. (2002) Using stochastic frontier analysis for performance measurement and benchmarking. *Econometric models in marketing*, Volume 16, pages 317-350.

[17] Stevenson, R.E. (1980). Likelihood function for generalized stochastic frontier estimation. *Journal of Econometrics,* 13,1,57-66.