STOCHASTIC FRONTIER ANALYSIS : A MATRIX REPRESENTATION

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

Comunicación Técnica No I-05-08/22-08-2005
(PE/CIMAT)

Stochastic Frontier Analysis: a matrix representation

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
Arjun K. Gupta
Bowling Green State University
Bowling Green, OH 43403-0221 USA
gupta@bgnet.bgsu.edu
and
Rogelio Ramos-Quiroga
Centro de Investigación en Matemáticas A. C.
Callejón de Jalisco s/n
36240 Guanajuato, MÉXICO
rramosq@cimat.mx

August 24, 2005

Abstract

Stochastic Frontier Analysis (*SFA*) models have been using skewness as an intrinsic characteristic to measure technical inefficiency. We extend the use of skew normality and elliptical errors in *SFA* as a flexible tool to model, for example, panel data. We consider stochastic frontier analysis in the common setting Normal + Truncated Normal with uncorrelated errors, as well as the case with correlated errors, in a matrix representation. The connection between the *SFA* model and the Closed Skew-Normal has been discussed in Domínguez-Molina, *et al* (2004). We provide a matrix representation for the skew-normal distribution and skew-elliptical distributions through a general setting and obtain conditional and marginal representations. Also, we obtain a useful submodel through an additive representation to be used with *SFA* models. We work the moment generating function and some quadratic forms of interest that allows several applications and in particular help to understand some properties in the *SFA* models.

Key Words: Stochastic frontier analysis, correlated error, random matrix, linear transformation, Wishart distribution, quadratic forms.

ABSTRACT

Stochastic Frontier Analysis (*SFA*) models have been using skewness as an intrinsic characteristic to measure technical inefficiency. We extend the use of skew normality and elliptical errors in *SFA* as a flexible tool to model, for example, panel data. We consider stochastic frontier analysis in the common setting Normal + Truncated Normal with uncorrelated errors, as well as the case with correlated errors, in a matrix representation. The connection between the *SFA* model and the Closed Skew-Normal has been discussed in Domínguez-Molina, *et al* (2004). We provide a matrix representation for the skew-normal distribution and skew-elliptical distributions through a general setting and obtain conditional and marginal representations. Also, we obtain a useful submodel through an additive representation to be used with *SFA* models. We work the moment generating function and some quadratic forms of interest that allow several applications and in particular help to understand some properties of the *SFA* models.

1. INTRODUCTION

The modeling of production functions to estimate productive efficiency of companies, had a breakthrough with the works of Aigner, Lovell and Schmidt (1977) and Meeusen and van den Broeck (1977); in these works, the concept of a stochastic frontier was introduced by letting

$$y = f(\mathbf{x}; \boldsymbol{\beta}) + \varepsilon \tag{1}$$

where the error term, $\varepsilon = v - u$, is composed by a symmetric disturbance term, v, representing measurement error, and by the nonnegative technical inefficiency component u. This formulation of the error structure recognizes that companies with the same technical ability to manage their resources might end up with different output levels (due to unobservable shocks v); the term u would be firm-specific and captures technical inefficiencies to reach the maximum theoretical output.

Assuming a cross-sectional data structure, Domínguez-Molina *et al* (2004) provide a stochastic frontier model based on the Closed Skew-Normal (CSN) distribution as given in González-Farías *et al* (2004a). Their proposal encompass nested submodels with increasing degree of complexity for the covariance structure, but within the framework of normal measurement errors and truncated normals for inefficiencies:

$$\mathbf{y} = \mathbf{f}(X; \boldsymbol{\beta}) + \mathbf{v} + G\mathbf{u}, \tag{2}$$

where $\mathbf{y}$ is the vector of value added observations on n firms, $\mathbf{f}$ is the production function commonly based on the Cobb-Douglas model with logged input variables; $\mathbf{v} \sim N_n(\mathbf{0}, \Sigma)$ models measurement error, and $\mathbf{u} \sim N_m^{\mathbf{c}}(\boldsymbol{\nu}, \Lambda)$, $m \geq n$, ($N_m^{\mathbf{c}}(\boldsymbol{\nu}, \Lambda)$ denotes the truncated $N_m(\boldsymbol{\nu}, \Lambda)$ at $\mathbf{c}$). This random vector $\mathbf{u}$ models shared technological inefficiencies in groups of firms, weighted by the $n \times m$ full row rank matrix G. Also, it is assumed that $\mathbf{v}$ is independent of $\mathbf{u}$, $\mathbf{f}(\mathbf{x}; \boldsymbol{\beta}) = (f(\mathbf{x}_1; \boldsymbol{\beta}), ..., f(\mathbf{x}_n; \boldsymbol{\beta}))'$, $X = (\mathbf{x}_1, \cdots, \mathbf{x}_n)'$ is a known matrix of covariates and $\boldsymbol{\beta}$ is unknown. The matrix G gives flexibility to the model: If it is left unspecified, it can be estimated and used to validate model assumptions, on the other hand, it can be defined as $G = I_n$ or $G = -I_n$ for firm-specific cost efficiencies or technical inefficiencies, respectively.

In this article we present an extension of the CSN distribution to the matrix-variate case having in mind panel data structures for the analysis of stochastic frontier model. The CSN distribution is closed under operations of marginalization and conditioning which are basic for statistical modeling. This class of distributions includes the normal distribution and has some properties like the normal family and yet they are skewed. In particular the expressions for the marginal and conditional densities are similar to those in the normal case. Several proposals of multivariate skew-normal distributions are special cases of the *CSN* distribution, thus, taking this distribution from the vector to the matrix case seems a useful and natural generalization. The matrix-variate *CSN* distribution that we propose, implicitly defines matrix-variate generalizations of several other multivariate skew-normal in the literature. The relationship between the closed skew normal and *SFA* models will be extended for the matrix case in the same way as it was done for the vector case in Domínguez-Molina *et al* (2004). A good account on skew distributions, either for vector or matrix representations on normal or elliptical distributions, theory and applications, can be found in Genton (2004).

The organization of this paper is as follows: In Section 2 we give the general setting for *SFA* models and its relationship with skew normal distributions. Section 3 provides the technical results that justify the matrix representation. Section 4 provides the basic properties for the matrix-variate skew normal and a proposal for a skew elliptical matrix representation of the error structure in our models. Some additional general results, in the context of matrix-variate skew distributions are also given. In Section 5, we discuss the results in terms of the *SFA* model, their advantages and limitations as well as future research. The proofs of the results obtained in the paper are given in the Appendix.

2. THE STOCHASTIC FRONTIER MODEL AND SKEW DISTRIBUTIONS

Econometricians have been interested in the specification and estimation of a frontier production function for over 30 years. Recently this problem has taken a new face with the integrated new economies, like the European Economic Community, bringing new and more complex data to analyze and for which we need to think in new and more flexible models to capture, as much as possible, the complex structure of the information on hand.

The original formulation, (1), of Aigner *et al* (1977) for the stochastic frontier model is

$$y = f(\mathbf{x}; \boldsymbol{\beta}) + v - u$$

the error term $\varepsilon = v - u$, when v is assumed $N(0, \sigma^2)$ distributed independently of $u \sim N^0(0, \tau^2)$, can be seen to have Azzalini's skew-normal distribution with density

$$g(\varepsilon) = 2\frac{1}{\varsigma}\phi\left(\frac{\varepsilon}{\varsigma}\right)\Phi\left(-\frac{\lambda}{\varsigma}\varepsilon\right),$$

where $\varsigma = \sqrt{\tau^2 + \sigma^2}$, $\lambda = \tau/\sigma$, and $\phi(\cdot)$ and $\Phi(\cdot)$ denote the density function and the distribution function of a standard normal random variable, respectively.
If the setting under consideration involves minimal cost frontiers, then the usual device is to switch the sign of u: $\varepsilon = v + u$. By doing this we have that the cost observations would be above the minimum cost frontier. A direct generalization, given by Domínguez-Molina *et al* (2004), is to consider

$$\varepsilon = v + \alpha u(c),$$

where α is fixed, and $u(c)$ is a random variable truncated below at a positive constant c. This idea led to the proposal of model (2)

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

Under the assumptions stated in the introduction, it is shown (see Domínguez-Molina *et al* (2004)) that the density of the compound error $\boldsymbol{\varepsilon} = \mathbf{v} + G\mathbf{u}$ is

$$g(\boldsymbol{\varepsilon}) = \Phi_m^{-1}(\mathbf{0}; \mathbf{c} - \boldsymbol{\nu}, \Lambda)\, \phi_n(\boldsymbol{\varepsilon}; G\boldsymbol{\nu}, \Theta)\, \Phi_m\left[\Lambda G'\Theta^{-1}(\boldsymbol{\varepsilon} - G\boldsymbol{\nu}); \mathbf{c} - \boldsymbol{\nu}, \Upsilon\right],$$

where

$$\Theta = \Sigma + G\Lambda G' \quad \text{and } \Upsilon = \Lambda - \Lambda G'\Theta^{-1}G\Lambda,$$

thus $\boldsymbol{\varepsilon}$ has a closed skew-normal distribution

$$\boldsymbol{\varepsilon} \sim CSN_{n,m}\left(G\boldsymbol{\nu}, \boldsymbol{\Theta}, \Lambda G'\Theta^{-1}, \mathbf{c} - \boldsymbol{\nu}, \Upsilon\right).$$

Model (2) includes the following cases as submodels

- Model I: Homoscedastic and uncorrelated errors

$$G = \alpha I_n, \quad \Sigma = \sigma^2 I_n, \quad \Lambda = \tau^2 I_n.$$

- Model II: Heteroscedastic and uncorrelated errors. G, Σ and Λ diagonal matrices, any of them of the form

$$G = \operatorname{diag}(\alpha_1, \ldots, \alpha_n), \quad \Sigma = \operatorname{diag}\left(\sigma_1^2, \ldots, \sigma_n^2\right), \quad \Lambda = \operatorname{diag}\left(\tau_1^2, \ldots, \tau_n^2\right).$$

- Model III: Correlated errors. If any of the matrices G, Σ or Λ are non-diagonal we would have the case of correlated errors.

- Model IV: Multiple output model of Kumbhakar & Lovell (2000, eq. 5.3.7).
  $\mathbf{v} \sim N_n(\mathbf{0}, \Sigma)$ and $u \sim N^0(0, \sigma^2)$, $G = \mathbf{1}_n\gamma$, $\gamma < 0$, $\mathbf{1}_n$ is a $n \times 1$ vector of ones.

The data structure for $\mathbf{y} = (y_1, y_2, \ldots, y_n)'$ in model (2) is assumed that of a cross sectional sample of n firms. We now consider panel data structures of the form

$$Y_{n\times m} = (\mathbf{y}_1, \ldots, \mathbf{y}_m),$$

where n firms are followed through times $t = 1, \cdots, m$. We are interested in the proposal of a multivariate stochastic frontier model that, in principle, can be described with the same type of error structure and with the ability to accommodate, in a simple way, as many as correlation patterns described before. Thus, we need a multivariate stochastic model of the form

$$Y = f(X; \boldsymbol{\beta}) + \Xi,$$

where

$$\Xi = V + DUE'.$$

In order to characterize this model, we need to develop the matrix variate closed skew normal representation and study its basic properties.

3. THE MULTIVARIATE AND MATRIX VARIATE CLOSED SKEW NORMAL DISTRIBUTIONS

The use of matrix theory is now extensive in both physical and social sciences. In order to have a more plausible stochastic frontier model we described the need of developing a matrix variate theory when we have truncation or, equivalently, a mechanism that induces a bias in the distribution.

Let X be an $p \times m$ matrix. We express it in terms of elements, columns, and rows as

$$X = (\mathbf{x}_1, ..., \mathbf{x}_m) = \left(\mathbf{x}_1^*, ..., \mathbf{x}_p^*\right)'.$$

Here $\mathbf{x}_1, ..., \mathbf{x}_m$ can be thought of as a sample of size m from a $p-$dimensional population, but it is not necessary that $\mathbf{x}_1, \cdots, \mathbf{x}_m$ be independent. The study of random matrices is the base of *sampling theory* in multivariate analysis.

There are several works in which authors have defined and studied many classes of multivariate skew-normal distributions (*MSN*), see Genton (2004), just to mention one. It is important to develop the *sample theory* based on *MSN* random vectors. In order to do this, we need to extend the concepts of *MSN* from the vector case to the matrix case. In this section we define a class of matrix variate closed skew normal (*MVCSN*) distribution. To do so, we need to reestablish some properties of the vector closed skew normal distribution that will give us more flexibility to generalize it for the matrix case.

Definition 1 *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 density function of the CSN distribution is given by:*

$$g_{p,q}(\mathbf{y}) = C\phi_p(\mathbf{y}; \boldsymbol{\mu}, \Sigma)\, \Phi_q[D(\mathbf{y} - \boldsymbol{\mu}); \boldsymbol{\nu}, \Delta], \quad \mathbf{y} \in \mathbb{R}^p, \tag{3}$$

with:

$$C^{-1} = \Phi_q(\mathbf{0}; \boldsymbol{\nu}, \Delta + D\Sigma D'), \tag{4}$$

where $\phi_p(\cdot; \boldsymbol{\eta}, \Psi)$ and $\Phi_p(\cdot; \boldsymbol{\eta}, \Psi)$ are the pdf and the cdf of a p-dimensional normal distribution, repeatedly. Here $\boldsymbol{\eta} \in \mathbb{R}^p$ denotes the mean and Ψ is a $p \times p$ covariance matrix.

We will denote a p-dimensional random vector distributed according to a *CSN* distribution with parameters $q, \boldsymbol{\mu}, \Sigma, D, \boldsymbol{\nu}, \Delta$ by $\mathbf{y} \sim CSN_{p,q}(\boldsymbol{\mu}, \Sigma, D, \boldsymbol{\nu}, \Delta)$.

We introduce here a useful result that allows us to find moments and to prove other propositions, the *Moment Generating Function*, given in Gonzalez-Farías *et al* (2004b)

$$M_{\mathbf{y}}(\mathbf{s}) = \frac{\Phi_q(D\Sigma\mathbf{s}; \boldsymbol{\nu}, \Delta + D\Sigma D')}{\Phi_q(\mathbf{0}; \boldsymbol{\nu}; \Delta + D\Sigma D')} e^{\mathbf{s}'\boldsymbol{\mu} + \frac{1}{2}\mathbf{s}'\Sigma\mathbf{s}}, \quad \mathbf{s} \in \mathbb{R}^p.$$

3.1. SOME PROPERTIES OF THE *CSN* DISTRIBUTION

Proposition 2 *Marginal Representation. Let $\mathbf{v} \sim N_p(\mathbf{0}, I_p)$, $\mathbf{u} \sim N_q^{\boldsymbol{\nu}}(\mathbf{0}, \Delta + D\Sigma D')$, $\mathbf{u}$ independent of $\mathbf{v}$. Then the distribution of*

$$\mathbf{y} = \boldsymbol{\mu} + \left(\Sigma^{-1} + D'\Delta^{-1}D\right)^{-1/2} \mathbf{v} + \Sigma D' (\Delta + D\Sigma D')^{-1} \mathbf{u},$$

is $CSN_{p,q}(\boldsymbol{\mu}, \Sigma, D, \boldsymbol{\nu}, \Delta)$.

Another way to motivate the closed skew density is through a hidden truncation process, in many applications this would be the reason why we finish with an asymmetric distribution of this kind. On the other hand, it also gives a useful way to establish some of its properties providing a better insight on the behavior for these skew distributions.

For completeness we include it here. *Conditional Representation:* Let $\mathbf{e}_1 \sim N_p(\mathbf{0}, \Sigma)$ and $\mathbf{e}_2 \sim N_q(\mathbf{0}, \Delta)$ be independent random vectors. Consider the transformed variables

$$\begin{aligned} \mathbf{w}_0 &= \boldsymbol{\mu} + \mathbf{e}_1 \\ \mathbf{z} &= -\boldsymbol{\nu} + D\mathbf{e}_1 + \mathbf{e}_2, \end{aligned} \tag{5}$$

where $D\,(q \times p)$ is an arbitrary matrix, $\boldsymbol{\mu} \in \mathbb{R}^p$, $\boldsymbol{\nu} \in \mathbb{R}^q$ and $\Delta\,(q \times q) > 0$.
Then the distribution of $\mathbf{w} \overset{\mathbf{d}}{=} \mathbf{w}_0 | \{\mathbf{z} \geq \mathbf{0}\}$ is $CSN_{p,q}(\boldsymbol{\mu}, \Sigma, D, \boldsymbol{\nu}, \Delta)$, where $\overset{d}{=}$ means equal in distribution.

Proposition 3 *The distribution function of a CSN random vector* $\mathbf{y}$*, with parameters* $\boldsymbol{\mu}, \Sigma,$ $D, \boldsymbol{\nu}, \Delta$ *is given by*

$$F_{p,q}(\mathbf{y}_0; \boldsymbol{\mu}, \Sigma, D, \boldsymbol{\nu}, \Delta) = C\Phi_{p+q}\left[\begin{pmatrix} \mathbf{y}_0 \\ \mathbf{0} \end{pmatrix}; \begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{\nu} \end{pmatrix}, \begin{pmatrix} \Sigma & -\Sigma D' \\ -D\Sigma & \Delta + D\Sigma D' \end{pmatrix}\right],$$

where C *is given in (4).*

A *conditional argument* that allows the use of some basic properties of the multivariate normal distribution (instead of having to show most of the results from its definition or through the *mgf*) is based on the following reasoning: Let $\mathbf{w}$ and $\mathbf{z}$ be two random vectors and consider the random vector $\mathbf{y}$ constructed as

$$\mathbf{y} = \mathbf{w} | \{\mathbf{z} \in B\},$$

then

$$h(\mathbf{y}) \overset{d}{=} h(\mathbf{w}) | \{\mathbf{z} \in B\}.$$

for any Borel function h. If we consider a collection of pairs of random vectors $\mathbf{w}_i$ and $\mathbf{z}_i$ and construct the collection $\mathbf{y}_1, \cdots, \mathbf{y}_n$, through $\mathbf{y}_i = \mathbf{w}_i | \{\mathbf{z}_i \in B_i\}$, $i = 1, 2, \cdots, n$. Then it follows that

$$h(\mathbf{y}_1, \cdots, \mathbf{y}_n) \overset{d}{=} h(\mathbf{w}_1, \cdots, \mathbf{w}_n) \,|\, \left\{\bigcap_{i=1}^{n} \{\mathbf{z}_i \in B_i\}\right\},$$

where h is any measurable function. For example if $\mathbf{w}_1, \cdots, \mathbf{w}_n$ are of the same dimension then

$$\mathbf{x}_1 + \cdots + \mathbf{x}_n \overset{d}{=} (\mathbf{w}_1 + \cdots + \mathbf{w}_n) \,|\, \left\{\bigcap_{i=1}^{n} \{\mathbf{z}_i \in B_i\}\right\}.$$

Díaz-García and González-Farías (2005) obtain the expression for the distribution of a linear transformation of a skew-normal random vector without any rank restrictions, showing that it is a closed operation. Full row rank and full column rank linear transformations were studied in González Farías *et al* (2004). That is, if A is an $n \times p$ matrix, then

$$A\mathbf{y} \sim CSN_{n,q}\left(\boldsymbol{\mu}_A,\ \Sigma_A,\ D_A,\ \boldsymbol{\nu},\ \Delta_A\right), \tag{6}$$

for an appropriate set of parameters $\boldsymbol{\mu}_A$, Σ_A, D_A and Δ_A, according to the rank of A.
As a further example, lets prove the result: Let $\mathbf{y} \sim CSN_{p,q}(\boldsymbol{\mu}, \Sigma, D, \boldsymbol{\nu}, \Delta)$ and A be an $r \times p$ $(r \leq p)$ constant matrix, such that AA' is a nonsingular matrix and let $\mathbf{b} \in \mathbb{R}^r$. Then

$$A\mathbf{y} + \mathbf{b} \sim CSN_{r,q}\left(A\boldsymbol{\mu} + \mathbf{b},\ \Sigma_A,\ D_A,\ \boldsymbol{\nu},\ \Delta_A\right),$$

where

$$\Sigma_A = A\Sigma A',\ D_A = D\Sigma A'\left(A\Sigma A'\right)^{-1},$$

and

$$\Delta_A = \Delta + D\Sigma D' - D\Sigma A'\left(A\Sigma A'\right)^{-1} A\Sigma D',$$

by applying the conditional argument as follows:

Given that $\mathbf{y} \sim CSN_{p,q}(\boldsymbol{\mu}, \Sigma, D, \boldsymbol{\nu}, \Delta)$, we get that $\mathbf{y} \stackrel{d}{=} \mathbf{w} | \{\mathbf{z} \geq \mathbf{0}\}$, where the joint distribution of $\mathbf{w}$ and $\mathbf{z}$ is given by (5). Thus $A\mathbf{y} + \mathbf{b} \stackrel{d}{=} (A\mathbf{w} + \mathbf{b}) \,|\, \{\mathbf{z} \geq \mathbf{0}\}$ and

$$\begin{pmatrix} A\mathbf{w} + \mathbf{b} \\ \mathbf{z} \end{pmatrix} \sim N_{q+p}\left[\begin{pmatrix} A\boldsymbol{\mu} + \mathbf{b} \\ -\boldsymbol{\nu} \end{pmatrix}, \begin{pmatrix} A\Sigma A' & A\Sigma D' \\ D\Sigma A' & \Delta + D\Sigma D' \end{pmatrix}\right],$$

given that

$$\begin{pmatrix} \Sigma & A\Sigma D' \\ D\Sigma A' & \Delta + D\Sigma D' \end{pmatrix} = \begin{pmatrix} \Sigma_A & \Sigma_A D_A' \\ D_A\Sigma & \Delta_A + D_A\Sigma_A D_A' \end{pmatrix},$$

and the result follows. ■

Remark 4 *If A is nonsingular $D_A = DA^{-1}$, which implies that $\Delta_A = \Delta$. That is, if $|A| \neq 0$ the parameter Δ is not affected by the linear transformation. Also notice that $\boldsymbol{\nu}$ is not affected by any matrix A.*

The following proposition provides some common results for quadratic forms that could be useful for hypothesis testing.

Proposition 5 *Let* $\mathbf{y}$ *be a* $CSN_{p,q}(\boldsymbol{\mu},\Sigma,D,\boldsymbol{\nu},\Delta)$ *random vector. Then for an arbitrary constant symmetric matrix* $A\,(p\times p)$ *the mgf of* $\mathbf{y}'A\mathbf{y}$ *is*

$$\begin{aligned}M_{\mathbf{y}'A\mathbf{y}}(t) &= \frac{\Phi_q\left\{D\left[(I-2A\Sigma t)^{-1}-I\right]\boldsymbol{\mu},\boldsymbol{\nu},\Delta+D\,(I-2A\Sigma t)^{-1}D'\right\}}{\Phi_q\left(0;\boldsymbol{\nu},\Delta+D\Sigma D'\right)\exp\left\{-\frac{1}{2}\boldsymbol{\mu}'\left[(I-2A\Sigma t)^{-1}-I\right]\Sigma^{-1}\boldsymbol{\mu}\right\}}\\ &\quad\times\left|I-2A\Sigma t\right|^{-1/2},\end{aligned}$$

where $t\in\mathbb{R}$.

Remark 6 *Let* $\Sigma=\Gamma\Lambda\Gamma'$ *be the spectral decomposition of* Σ*, where* Λ *is a diagonal matrix of eigenvalues of* Σ *and* Γ *is an orthogonal matrix whose columns are normalized eigenvectors of* Σ*. Then, for* $D=\Gamma$*,* $\boldsymbol{\mu}=\mathbf{0}$*,* $\boldsymbol{\nu}=\mathbf{0}$ *and* Δ *diagonal,from Proposition 5 we get*

$$\begin{aligned}M_{\mathbf{y}'\Sigma^{-1}\mathbf{y}}(t) &= \frac{\Phi_q\left[\mathbf{0},\mathbf{0},\Delta+\Gamma\left(I_p-2\Sigma^{-1}\Sigma t\right)^{-1}\Gamma'\right]}{\Phi_q\left(\mathbf{0};\mathbf{0},\Delta+\Gamma\Sigma\Gamma'\right)}\left|I_p-2t\Sigma\Sigma^{-1}\right|^{-1/2}\\ &= \frac{\Phi_q\left[\mathbf{0},\mathbf{0},\Delta+\left(\Gamma\Gamma'-2\Gamma\Gamma' t\right)^{-1}\right]}{\Phi_q\left(\mathbf{0};\mathbf{0},\Delta+\Lambda\right)}\left|I_p-2tI_p\right|^{-1/2}\\ &= \left|I_p-2tI_p\right|^{-1/2}\\ &= (1-2t)^{-p/2}.\end{aligned}$$

Thus

$$\mathbf{y}'\Sigma^{-1}\mathbf{y}\sim\chi^2_p,$$

where χ^2_p *denotes the chi-square distribution with* p *degrees of freedom.*

In fact, if $D\Sigma D'$ is an arbitrary diagonal matrix under the same restrictions for $\boldsymbol{\mu},\boldsymbol{\nu}$ and Δ we get the same result.

Remark 7 *The joint mgf of* $\mathbf{y}'A_1\mathbf{y}$ *and* $\mathbf{y}'A_2\mathbf{y}$ *for any two arbitrary* $p\times p$ *matrices* A_1 *and* A_2 *is obtained by noting that*

$$\begin{aligned}M_{\mathbf{y}'A_1\mathbf{y},\mathbf{y}'A_2\mathbf{y}}(t_1,t_2) &= Ee^{t_1\mathbf{y}'A_1\mathbf{y}+t_2\mathbf{y}'A_2\mathbf{y}}\\ &= Ee^{\mathbf{y}'(t_1A_1+t_2A_2)\mathbf{y}}\\ &= M_{\mathbf{y}'(t_1A_1+t_2A_2)\mathbf{y}}(1).\end{aligned}$$

Some of the most important properties of the CSN distributions are their closure properties. For example, the joint distribution of independent CSN variables belongs to the same family as well as the sum of independent CSN random variables. This closure property makes the study of distributional properties of random samples more tractable, and will be very useful for the extension to the matrix variate case under certain types of dependencies. We give the following two results and establish a quick proof for the sum, based on the conditional argument.

The symbols " $\otimes$ 」 and " $\oplus$ 」 are the Kronecker matrix-product and the matrix direct sum operator (see Horn & Johnson, 1985, page 24): For any two matrices A and B, $A \otimes B = (a_{ij}B)$, $A \oplus B$ gives as a result a block diagonal matrix. Read $\bigoplus_{i=1}^{m} A_i$ as $A_1 \oplus \cdots \oplus A_m$, for any matrices $A_1, \cdots, A_m$. Here $\mathbf{I}_m$ is the $m \times m$ identity matrix. We denote by $\mathbf{1}_m$ the $(m \times 1)$ vector of ones, that is $\mathbf{1}'_m = (1, \cdots, 1) = \sum_{i=1}^{m} e'_{m,i}$, where $e_{m,i} = (0, \cdots, 0, 1, 0, \cdots, 0)'$ is the $m \times 1$ vector with 1 at the i-th position.

Let $\mathbf{y}_i \, (p_i \times 1)$ and

$$\mathbf{u}_i = \left(\mathbf{y}'_i, \ \mathbf{x}'_{iq_i}\right)' \sim N_{p_i+q_i}\left[\begin{pmatrix} \mu_i \\ -\nu_i \end{pmatrix}, \begin{pmatrix} \Sigma_i & \Sigma_i D'_i \\ D_i\Sigma_i & \Delta_i + D_i \Sigma D' \end{pmatrix}\right].$$

Let $\mathbf{u} = (\mathbf{u}'_1, \cdots, \mathbf{u}'_n)'$, and consider $Q = \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix}$ such that

$$Q\mathbf{u} = \left(\mathbf{y}'_1, \mathbf{y}'_2, \cdots, \mathbf{y}'_n, \mathbf{x}'_{1q_1}, \cdots, \mathbf{x}'_{nq_n}\right)',$$

where Q_1 and Q_2 are

$$Q_1 = (Q'_{11}, \cdots, Q'_{1n})', \quad Q_2 = (Q'_{21}, \cdots, Q'_{2n})',$$

and

$$Q_{1i} = \left(0_{p_i \times r_{i-1}} \ I_{p_i} \ 0_{p_i \times (r_n - r_{i-1} - p_i)}\right), \quad Q_{2i} = \left(0_{q_i \times (r_i - q_i)} \ \ I_{q_i} \ \ 0_{q_i \times (r_n - r_i)}\right),$$

$$r_i = \sum_{k=1}^{i} (p_k + q_k), \quad i = 1, 2, ..., n.$$

Observe that

$$\mathbf{y} = (\mathbf{y}'_1, \cdots, \mathbf{y}'_n)' = Q_1 \mathbf{u},$$

and

$$\mathbf{x} = \left(\mathbf{x}'_{1q_1}, \cdots, \mathbf{x}'_{nq_n}\right) = Q_2\mathbf{u}.$$

Thus, the distribution of $\mathbf{y}|\left\{\mathbf{x} \geq 0\right\}$ is given by the following theorem.

Proposition 8 *If $\mathbf{y}_1, \cdots, \mathbf{y}_n$ are independent random vectors from the $CSN_{p_i,q_i}(\boldsymbol{\mu}_i, \Sigma_i, D_i, \boldsymbol{\nu}_i, \Delta_i)$ distribution, then the joint distribution of $\mathbf{y}_1, \cdots, \mathbf{y}_n$ is*

$$\mathbf{y} = (\mathbf{y}'_1, ..., \mathbf{y}'_n)' \sim CSN_{p^\dagger, q^\dagger}\left(\boldsymbol{\mu}^\dagger, \Sigma^\dagger, D^\dagger, \boldsymbol{\nu}^\dagger, \Delta^\dagger\right),$$

where $p^\dagger = \sum_{i=1}^n p_i$, $\quad q^\dagger = \sum_{i=1}^n q_i$, $\quad \boldsymbol{\mu}^\dagger = (\boldsymbol{\mu}'_1, \cdots, \boldsymbol{\mu}'_n)'$, $\ \Sigma^\dagger = \bigoplus_{i=1}^n \Sigma_i$, $\ D^\dagger = \bigoplus_{i=1}^n D_i$, $\quad \boldsymbol{\nu}^\dagger = (\boldsymbol{\nu}'_1, \cdots, \boldsymbol{\nu}'_n)'$, $\quad \Delta^\dagger = \bigoplus_{i=1}^n \Delta_i$.

In order to obtain the distribution of the sum of independent CSN random vectors we must take $p_i = p$, $i = 1, \cdots, n$.

For $A = \left(\mathbf{1}'_n \otimes \mathbf{I}_p \ \ 0_{p\times(r_n - np)}\right) Q_1$, we have

$$\sum_{i=1}^n \mathbf{y}_i = A\mathbf{u},$$

and by computing the conditional density of

$$(A\mathbf{u})\,|\,\{\mathbf{x} \geq 0\},$$

we obtain the following Proposition.

Proposition 9 *If $\mathbf{y}_1, \cdots, \mathbf{y}_n$ are independent random vectors with $\mathbf{y}_i \sim CSN_{p,q_i}(\boldsymbol{\mu}_i, \Sigma_i, D_i, \boldsymbol{\nu}_i, \Delta_i)$, $i = 1, 2, \cdots, n$, then*

$$\sum_{i=1}^n \mathbf{y}_i \sim CSN_{p,q^\star}\left(\boldsymbol{\mu}^\star, \Sigma^\star, D^\star, \boldsymbol{\nu}^\star, \Delta^\star\right),$$

where $q^\star = \sum_{i=1}^n q_i$, $\quad \boldsymbol{\mu}^\star = \sum_{i=1}^n \boldsymbol{\mu}_i$, $\quad \Sigma^\star = \sum_{i=1}^n \Sigma_i$,

$$D^\star = (\Sigma_1 D'_1, ..., \Sigma_n D'_n)' \left(\sum_{i=1}^n \Sigma_i\right)^{-1}, \quad \boldsymbol{\nu}^\star = (\boldsymbol{\nu}'_1, ..., \boldsymbol{\nu}'_n)',$$

and

$$\Delta^\star = \Delta^\dagger + D^\dagger \Sigma^\dagger D^{\dagger\prime} - D^\star \left(\sum_{i=1}^n \Sigma_i\right) D^{\star\prime}.$$

3.2. THE EXTENDED SKEW ELLIPTICAL DISTRIBUTION

González-Farías, *et al* (2004) obtained a multivariate extended skew-elliptical (ESE) distribution in a similar way as the CSN distribution.

We consider elliptical random vectors whose density function exists and $P(\mathbf{y}=\mathbf{0})=0$. Suppose that a $p-$dimensional random vector $\mathbf{y}$ has a density of the form:

$$f_p\left(\mathbf{y};\boldsymbol{\eta},\Theta;h\right)=\left|\Theta\right|^{-1/2}h\left[\left(\mathbf{y}-\boldsymbol{\eta}\right)'\Theta^{-1}\left(\mathbf{y}-\boldsymbol{\eta}\right)\right], \tag{7}$$

where $\boldsymbol{\eta}\in\mathbb{R}^p$, Θ is a $p\times p$ positive definite matrix and $h\left(\cdot\right)$ is a nonnegative function of a scalar variable such that

$$\pi^{\frac{p}{2}}\int_0^\infty t^{\frac{p}{2}-1}h\left(t\right)dt=\Gamma\left(\tfrac{p}{2}\right).$$

If $\mathbf{y}$ has density function given by (7) we say that $\mathbf{y}$ has an elliptically contoured distribution and it will be denoted by $\mathbf{y}\sim EC_p\left(\boldsymbol{\eta},\Theta,h\right)$; and $h\left(\cdot\right)$ is called the *pdf* generator of the elliptic distribution. This is a rich family that contains, for example, heavy tail distributions as the t, Cauchy and so forth.

Let

$$\begin{pmatrix}\mathbf{w}_0\\ \mathbf{z}\end{pmatrix}\sim EC_{q+p}\left[\begin{pmatrix}\boldsymbol{\mu}\\ -\boldsymbol{\nu}\end{pmatrix},\begin{pmatrix}\Sigma & \Sigma D'\\ D\Sigma & \Delta+D\Sigma D'\end{pmatrix};h\right]. \tag{8}$$

where $\mathbf{w}_0$ is a random $p-$vector, $\mathbf{z}$ is a random $q-$vector, $\boldsymbol{\mu}\in\mathbb{R}^p,\boldsymbol{\nu}\in\mathbb{R}^q,$ D is an arbitrary $q\times p$ matrix, and Σ and Δ are $p\times p$ and $q\times q$ positive definite matrices, respectively.

Let h_2 be the *pdf* generator of the marginal distribution of $\mathbf{z}$. From Theorem 2.18 of Fang *et al* (1990) the conditional density of $\mathbf{z}$ given $\mathbf{w}_0=\mathbf{w}$ is given by

$$\mathbf{z}|\left\{\mathbf{w}_0=\mathbf{w}\right\}\sim EC_q\left(-\boldsymbol{\nu}+D\left(\mathbf{w}-\boldsymbol{\mu}\right),\Delta;h_{s(\mathbf{w})}\right),$$

where

$$h_a\left(u\right)=\frac{h\left(a+u\right)}{h_2\left(a\right)},\ a,\ u>0. \tag{9}$$

It follows that if $\mathbf{w}\overset{d}{=}\mathbf{w}_0|\left\{\mathbf{z}\geq\mathbf{0}\right\}$ then the *pdf* of $\mathbf{w}$ is:

$$f\left(\mathbf{w}\right)=\frac{f_p\left(\mathbf{w};\boldsymbol{\mu},\Sigma;h_1\right)F_q\left[D\left(\mathbf{w}-\boldsymbol{\mu}\right);\boldsymbol{\nu},\Delta;h_{s(\mathbf{w})}\right]}{F_q\left(\mathbf{0};\boldsymbol{\nu},\Delta+D\Sigma D';h_2\right)}, \tag{10}$$

where $f_p\left(\cdot;\cdot,\cdot;h_1\right)$ is as in (7), $h_{s(\mathbf{w})}$ is given in (9) and $F_q\left(\cdot;\cdot,\cdot;h^\dagger\right)$ is the distribution function of a random vector with *pdf* generator $h^\dagger$.

We will denote a random p-vector, $\mathbf{y}$, distributed according to an *ESE* distribution with parameters $\boldsymbol{\mu}, \Sigma, D, \boldsymbol{\nu}, \Delta$ and *pdf* generator h by $\mathbf{y} \sim ESE_{p,q}(\boldsymbol{\mu}, \Sigma, D, \boldsymbol{\nu}, \Delta; h)$.
Using the closure properties of elliptical distributions for linear transformations in (8) it is easy to prove that the *ESE* distributions are closed under full row rank linear transformations.

3.3. THE MATRIX VARIATE *CSN* DISTRIBUTION

In the previous sections we defined and studied the properties of a *CSN* random vector that can be easily extended to an *ESE* random vector. The issue of how to extend the concept of a *CSN* or an *ESE* distribution from the vector case to the matrix case, is an attractive and important problem.
Define the observation random matrix as

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{p1} & \cdots & x_{pm} \end{pmatrix} = (\mathbf{x}_1, \cdots, \mathbf{x}_m) = \begin{pmatrix} \mathbf{x}_1^{*\prime} \\ \vdots \\ \mathbf{x}_p^{*\prime} \end{pmatrix},$$

where $\mathbf{x}_i\,(p \times 1)\,,\ i = 1, \cdots, m$ is the i^{th} column of X and $\mathbf{x}_i^{*\prime}$ is the i^{th} row of the matrix X. Here $\mathbf{x}_1, \cdots, \mathbf{x}_m$ can be thought as a sample of size m from a p-dimensional population, but it is not necessary that $\mathbf{x}_1, \cdots, \mathbf{x}_m$ are independent.

The random matrix $X\,(p \times m)$ is said to have a matrix variate normal distribution with mean matrix $M\,(p \times m)$ and covariance matrix $\Sigma \otimes \Psi$ where $\Sigma\,(p \times p) > 0$ and $\Psi\,(m \times m) > 0$, if $\operatorname{vec}(X') \sim N_{pm}(\operatorname{vec}(M'), \Sigma \otimes \Psi)$.
For a matrix $X\,(p \times m)$, $\operatorname{vec}(X)$ is the $pm \times 1$ vector defined as

$$\operatorname{vec}(X) = \begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_n \end{pmatrix}.$$

We will use the notation $X \sim N_{p,n}(M, \Sigma \otimes \Psi)$ and denote the *cdf* of X as

$$\Phi_{p,m}(X; \mu, \Omega) = \Phi_{pm}(\operatorname{vec}(X'); \operatorname{vec}(\mu'), \Omega).$$

Definition 10 *A random matrix $Y\,(p \times m)$ is said to have a matrix variate closed skew normal distribution with parameters*

$$M\,(p \times m)\,, S\,(pm \times pm)\,, B\,(pm \times qn)\,, L\,(q \times n)\,,\ Q\,(qn \times qn)\,,$$

$S>0$ *and* $Q>0,$ *if*

$$\operatorname{vec}(Y') \sim CSN_{pm,qn}\left[\operatorname{vec}(M'), S, B, \operatorname{vec}(L'), Q\right]. \tag{11}$$

We will use the notation

$$Y \sim CSN_{p,m;q,n}(M,S,B,L,Q). \tag{12}$$

For most of the cases the matrices S and B will have a specific structure. Most of the properties for the parametrization (12) are immediately obtained from González-Farías *et al* (2004).

Example. (*The distribution of the transpose of a CSN sample):*
Let $Y=(\mathbf{y}_1,\cdots,\mathbf{y}_n)$ be a sample of independent and identically distributed random vectors from the $CSN_{p,q}(\boldsymbol{\mu},\Sigma,D,\boldsymbol{\nu},\Delta)$ distribution. From González-Farías, *et al* (2004) we get that the distribution of

$$\operatorname{vec}(Y) = (\mathbf{y}_1',...,\mathbf{y}_n')',$$

is

$$CSN_{np,nq}(\mathbf{1}_n\otimes\boldsymbol{\mu}, I_n\otimes\Sigma, I_n\otimes D, \mathbf{1}_n\otimes\boldsymbol{\nu}, I_n\otimes\Delta),$$

and hence we obtain that

$$Y' \sim CSN_{p,n;q,n}(\mathbf{1}_n'\otimes\boldsymbol{\mu}, I_n\otimes\Sigma, I_n\otimes D, \mathbf{1}_n'\otimes\boldsymbol{\nu}, I_n\otimes\Delta).$$

The joint distribution of the *CSN* sample Y is given in Corollary 13.

3.3.1. A MATRIX MODEL

Let $U_1\sim N_{p,m}(0,S)$ and $U_2\sim N_{q,n}(0,Q)$ be independent normal random matrices.
Consider the model

$$\begin{aligned} W &= M+U_1 \\ Z &= -L+DU_1E'+U_2, \end{aligned}$$

$D(q\times p)$, $E(m\times n)$ then the joint distribution of W and Z is

$$\begin{pmatrix} W \\ Z \end{pmatrix} \sim N_{qn+pm}\left[\begin{pmatrix} M \\ -L \end{pmatrix}, \Omega\right],$$

with

$$\Omega = \begin{pmatrix} S & S(D'\otimes E') \\ (D\otimes E)S & Q+(D\otimes E)S(D'\otimes E') \end{pmatrix}.$$

Now, if

$$Y \stackrel{d}{=} W | \{Z \geq 0\},$$

we get that

$$f(Y) = K\phi_{p,m}(Y; M, S)\,\Phi_{q,n}\left[E(Y - M)D'; L, Q\right],$$

where

$$K^{-1} = \Phi_{q,n}\left[0; L, Q + (D \otimes E)S(D' \otimes E')\right],$$

hence

$$Y \sim CSN_{p,m;q,n}(M, S, D \otimes E, L, Q).$$

Which is a particular case of (12).

4. PROPERTIES OF THE MVCSN DISTRIBUTION

4.1. DISTRIBUTION OF THE TRANSPOSE

In order to compute the distribution of the transpose of Y we need to define the commutation matrix which transforms $\mathrm{vec}(A)$ into $\mathrm{vec}(A')$. The commutation matrix $K_{mp}\,(mp \times mp)$ is defined as

$$K_{mp} = \sum_{i=1}^{m}\sum_{j=1}^{p}\left(H_{ij} \times H'_{ij}\right),$$

where $H_{ij}\,(m \times p)$ has a unit element at the $(i,j)^{th}$ place and zero elsewhere. Thus if $Y \sim CSN_{p,m;q,n}(M, S, B, L, Q)$ the distribution of Y' can be obtained, from the fact that

$$\mathrm{vec}(Y) = K_{mp}\mathrm{vec}(Y'),$$

and then, by using (6), we get the following result

Proposition 11 *Let* $Y \sim CSN_{p,m;q,n}(M, S, B, L, Q)$ *then*

$$Y' \sim CSN_{m,p;n,q}(M', K_{mp}SK_{pm}, K_{pm}B', L, Q).$$

Corollary 12 *If in Proposition (11)* $S = \Sigma \otimes \Psi$ *then*

$$Y' \sim CSN_{m,p;n,q}(M', \Psi \otimes \Sigma, K_{pm}B', L, Q).$$

Corollary 13 *If* $\mathbf{y}_1, \cdots, \mathbf{y}_n$ *are independent and identically distributed random vectors from the* $CSN_{p,q}(\boldsymbol{\mu}, \Sigma, D, \boldsymbol{\nu}, \Delta)$ *distribution, then the joint distribution of* $Y = (\mathbf{y}_1, \cdots, \mathbf{y}_n)$ *is:*

$$Y \sim CSN_{n,p;n,q}\left(\mathbf{1}_n \otimes \boldsymbol{\mu}', \Sigma \otimes I_n, (D' \otimes I_n)K_{nq}, \mathbf{1}'_n \otimes \boldsymbol{\nu}, I_n \otimes \Delta\right).$$

4.2. LINEAR TRANSFORMATIONS

Proposition 14 *Consider $Y \sim CSN_{p,m;q,n}(M,S,B,L,Q)$ and let $A_1(n_1 \times p)$, $A_2(m \times n_2)$ matrices such that $A = A_1 \otimes A_2'$ has full row rank. If $W = A_1 Y A_2$ then*

$$W \sim CSN_{p,m;q,n}\left(M_A,\ S_A,\ B_A,\ L,\ Q_A\right),$$

where

$$M_A = A_1 M A_2, \quad S_A = ASA', \quad B_A = BSA'S_A^{-1},$$

and

$$Q_A = Q + BSB' - BSA'S_A^{-1}ASB'.$$

4.3. MOMENT GENERATING FUNCTION

Lemma 15 *Let $W \sim N_{p,m}(M,S)$, where $M(p \times m)$ and $\Sigma(p \times p) > 0$, $\Psi(m \times m) > 0$. Then*

$$E_W\left[\Phi_{q,n}(A + BWC; L, Q)\right] = \Phi_{q,n}\left[A + BMC; L, Q + (C' \otimes B) S (C \otimes B')\right],$$

where $A(q \times n)$, $B(q \times p)$, $C(m \times n)$, $L(q \times n)$, $\Lambda(q \times q)$, $\Upsilon(n \times n)$; $\Lambda > 0$, $\Upsilon > 0$.

Proposition 16 *Let $Y \sim CSN_{p,m;q,n}(M, \Sigma \otimes \Psi, D \otimes E, L, Q)$. Then the mgf of Y is given by*

$$\begin{aligned} M_Y(T) &= E\operatorname{etr}(Y'T) \\ &= \frac{\Phi_{q,n}\left(E\Psi T'\Sigma D'; L, Q + (D\Sigma D') \otimes (E\Psi E')\right)}{\Phi_{q,n}\left[0; L, Q + (D\Sigma D') \otimes (E\Psi E')\right]} \\ &\times \operatorname{etr}\left(M'T + \tfrac{1}{2}T'\Sigma T\Psi\right). \end{aligned}$$

4.4. QUADRATIC FORMS

Proposition 17 *Let $A(r \times m)$, $B(p \times p)$, $C(m \times s)$, $r \leq m$, $s \leq m$, and*

$$Y \sim CSN_{p,m;q,n}(0, \Sigma \otimes \Psi, D \otimes E, L, Q).$$

Then the mgf of

$$Z = AY'BYC$$

is

$$M_Z(T)=\frac{\Phi_{qn}\left[0,L,Q+(D\otimes E)\Theta(D'\otimes E')\right]}{\Phi_{qn}\left[0,L,Q+(D\Sigma D')\otimes(E\Psi E')\right]}\left|I_{pm}-2(\Sigma B)\otimes(\Psi CT'A)\right|^{-1/2}, \qquad (13)$$

where $\Theta=\left[I-2(B\Sigma)\otimes(CT'A\Psi)\right]^{-1}$.

Corollary 18 *Let* $Y\sim CSN_{p,m;1,1}(0,\Sigma\otimes\Psi,D\otimes E,0,\vartheta)$, $A=C=I_m$, *then* $Y'\Sigma^{-1}Y$ *has Wishart distribution with parameters* m, q *and* Ψ, *that is*

$$Y'\Sigma^{-1}Y\sim W_m(p,\Psi).$$

Corollary 19 *Let* $Y\sim CSN_{p,1;p,n}(0,\Sigma,\Gamma\otimes E,0,Q)$, *where* Γ *is part of the spectral decomposition of* Σ, $\Sigma=\Gamma\Lambda\Gamma'$ *and* Q *is diagonal. Then*

$$Y'\Sigma^{-1}Y\sim\chi^2_p.$$

4.5. THE MATRIX VARIATE EXTENDED SKEW-ELLIPTICAL DISTRIBUTION

A random matrix $Y(p\times m)$ is said to have a matrix variate extended skew-elliptical distribution with parameters

$$M(p\times m), S(pm\times pm), B(pm\times qn), L(q\times n),\ Q(qn\times qn)$$

$S>0$ and $Q>0$ with *pdf* generator h. If

$$\operatorname{vec}(Y')\sim ESE_{pm,qn}\left[\operatorname{vec}(M'),S,B,\operatorname{vec}(L'),Q,h\right].$$

We will use the notation

$$Y\sim ESE_{p,m;q,n}(M,S,B,L,Q,h).$$

5. A MULTIVARIATE STOCHASTIC FRONTIER MODEL

We will use the notation $U\sim N^{\mathbf{c}}_{m,n}(M,S)$ to denote a truncated $N_{m,n}(M,S)$ random matrix below at C, that is, the truncation is of the type $U\geq C$, where $W\geq C$ means $W_{ij}\geq C_{ij}$, $i=1,\cdots,m$, $j=1,\cdots,n$. Observe that $U\geq C\Rightarrow\operatorname{vec}(U')\geq\operatorname{vec}(C')$.
Consider production data on m firms at time t. Let us assume a stochastic frontier model for time t of the form

$$\mathbf{y}_t=\mathbf{f}(X_t,\boldsymbol{\beta}_t)+\boldsymbol{\varepsilon}_t,$$

where $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \cdots, \varepsilon_{mt})'$ is the random vector of compound errors

$$\boldsymbol{\varepsilon}_t = \mathbf{v}_t + D\mathbf{u}_t,$$

with $\mathbf{v}_t = (v_{1t}, \cdots, v_{mt})'$, $\mathbf{u}_t = (u_{1t}, \cdots, u_{qt})'$ and D a $m \times q$ weighting matrix. Let Y be the $m \times n$ matrix of all the value added observations on the m firms at times $t = 1, \cdots, n$,

$$Y = \begin{pmatrix} y_{11} & \cdots & y_{1n} \\ \vdots & \ddots & \vdots \\ y_{m1} & \cdots & y_{mn} \end{pmatrix} = (\mathbf{y}_1, \cdots, \mathbf{y}_n).$$

A joint model for the production data can be written as

$$Y = F + \Xi,$$

where $F = (\mathbf{f}(X_1, \boldsymbol{\beta}_1), \cdots, \mathbf{f}(X_n, \boldsymbol{\beta}_n))$, $\Xi = V + DU$, $V = (\mathbf{v}_1, \cdots, \mathbf{v}_n)$ and $U = (\mathbf{u}_1, \cdots, \mathbf{u}_n)$. We can consider a slightly more general model for the compound error

$$\Xi = V + DUE',$$

where $V \sim N_{m,n}(\mathbf{0}, S)$, $U \sim N^{\mathbf{c}}_{p,q}(L, Q)$, $D\,(m \times p)$, $E\,(n \times q)$, and V independent of U. Given that

$$\mathrm{vec}\,(Y') = \mathrm{vec}\,(V') + (D \otimes E)\,\mathrm{vec}\,(U')$$

from Domínguez-Molina, *et al* (2004) we get that the density of the compound error $\Xi = V + DUE'$ is:

$$\begin{aligned} g(\Xi) = {} & \Phi^{-1}_{p,q}(0; C - L, Q)\, \phi_{m,n}(\Xi; (D \otimes E) L, \Theta) \\ & \times \Phi_{m,n}\left[Q\,(D' \otimes E')\,\Theta^{-1}(\Xi - (D \otimes E) L); C - L, \Upsilon\right], \end{aligned}$$

where

$$\Theta = S + (D \otimes E)\, Q\, (D' \otimes E') \quad \text{and}$$

$$\Upsilon = Q - Q\,(D' \otimes E')\,\Theta^{-1}(D \otimes E)\, Q.$$

Thus Ξ has a matrix variate closed skew-normal distribution, *i.e.*,

$$\Xi \sim CSN_{p,m;n,q}\left[(D \otimes E) L, \Theta, QD'\Theta^{-1}, C - L, \Upsilon\right].$$

as defined in Section 3.3. Notice that the matrix V, of measurement errors, no longer is constrained to reflect measurement error, but also, depending on the structure of its variance matrix S, it can incorporate random effects such as random intercepts and time induced correlations among the columns of Y. The matrix of technical inefficiencies, U, by being pre-multiplied by D can incorporate common inefficiencies inside groups of similar companies and, by being post-multiplied by E' it can consider time related inefficiencies effects.

In this paper we have emphasized the close relationship that Stochastic Frontier Analysis has with the skew distributions; in particular with the *CSN* distribution and its matrix extension the *MVCSN*. The computational issues related to maximum likelihood estimation will always be present in applications and, of course, the consideration of parsimonious models is always a safe recommendation.

Acknowledgements

This work was partially supported by the research project 45974-F CONACYT-México, SEJ2004-04101-ECON, Spain and by CONCYTEG Grant 04-16-K117-028.

6. APPENDIX

Proof of Proposition 2. In order to obtain distribution of $\mathbf{y}$ we will use the mgf technique:

$$\begin{aligned}
M_{\mathbf{y}}(\mathbf{s}) &= e^{\mathbf{s}'\boldsymbol{\mu}} M_{\mathbf{v}}\left[\left(\Sigma^{-1}+D'\Delta^{-1}D\right)^{-1/2}\mathbf{s}\right] M_{\mathbf{u}}\left[\left(\Delta+D\Sigma D'\right)^{-1}D\Sigma\mathbf{s}\right] \\
&= e^{\mathbf{s}'\boldsymbol{\mu}} e^{\frac{1}{2}\mathbf{s}'\left(\Sigma^{-1}+D'\Delta^{-1}D\right)^{-1}\mathbf{s}} e^{\frac{1}{2}\mathbf{s}'\Sigma D'(\Delta+D\Sigma D')^{-1}(\Delta+D\Sigma D')(\Delta+D\Sigma D')^{-1}D\Sigma\mathbf{s}} \\
&\times \frac{\Phi_q\left(D\Sigma\mathbf{s};\boldsymbol{\nu},\Delta+D\Sigma D'\right)}{\Phi_q\left(\mathbf{0};\boldsymbol{\nu},\Delta+D\Sigma D'\right)} \\
&= e^{\mathbf{s}'\boldsymbol{\mu}} e^{\frac{1}{2}\mathbf{s}'\left[\left(\Sigma^{-1}+D'\Delta^{-1}D\right)^{-1}+\Sigma D'(\Delta+D\Sigma D')^{-1}D\Sigma\right]\mathbf{s}} \\
&\times \frac{\Phi_q\left(D\Sigma\mathbf{s};\boldsymbol{\nu},\Delta+D\Sigma D'\right)}{\Phi_q\left(\mathbf{0};\boldsymbol{\nu},\Delta+D\Sigma D'\right)}.
\end{aligned}$$

Using the Sherman-Morrison-Woodbury formula, we get

$$\left(\Sigma^{-1}+D'\Delta^{-1}D\right)^{-1}+\Sigma D'\left(\Delta+D\Sigma D'\right)^{-1}D\Sigma=\Sigma$$

Thus

$$M_{\mathbf{y}}(\mathbf{s})=\frac{\Phi_q\left(D\Sigma\mathbf{s};\boldsymbol{\nu},\Delta+D\Sigma D'\right)}{\Phi_q\left(\mathbf{0};\boldsymbol{\nu},\Delta+D\Sigma D'\right)}e^{\mathbf{s}'\boldsymbol{\mu}+\frac{1}{2}\mathbf{s}'\Sigma\mathbf{s}},$$

which is the mgf of a $CSN_{p,q}\left(\boldsymbol{\mu},\Sigma,D,\boldsymbol{\nu},\Delta\right)$ random vector. ■

Proof of Proposition 3 We know that

$$F_{p,q}(\mathbf{y}_0;\boldsymbol{\mu},\Sigma,D,\boldsymbol{\nu},\Delta)=\Pr(\mathbf{Y}\leq\mathbf{y}_0),$$

now, from the model (5) we get that

$$\begin{aligned}\Pr(\mathbf{y}\leq\mathbf{y}_0)&=\Pr(\mathbf{w}_0\leq\mathbf{y}_0|\mathbf{z}\geq\mathbf{0})\\&=\frac{\Pr(\mathbf{w}_0\leq\mathbf{y}_0,\mathbf{z}\geq\mathbf{0})}{\Pr(\mathbf{z}\geq\mathbf{0})}\\&=\frac{\Pr(\mathbf{w}_0\leq\mathbf{y}_0,-\mathbf{z}\leq\mathbf{0})}{\Pr(-\mathbf{z}\leq\mathbf{0})}\\&=C\Pr(\mathbf{w}_0\leq\mathbf{y}_0,-\mathbf{z}\leq\mathbf{0}),\end{aligned}$$

the result follows by noting that

$$\begin{pmatrix}\mathbf{w}_0\\-\mathbf{z}\end{pmatrix}\sim\Phi_{p+q}\left[\begin{pmatrix}\boldsymbol{\mu}\\\boldsymbol{\nu}\end{pmatrix},\begin{pmatrix}\Sigma & -\Sigma D'\\-D\Sigma & \Delta+D\Sigma D'\end{pmatrix}\right].$$

■

Proof of Proposition 5. It follows by first using the identity

$$\begin{aligned}e^{\mathbf{y}'A\mathbf{y}}\phi_p(\mathbf{y};\boldsymbol{\mu},\Sigma)&=|I_p-2A\Sigma t|^{-1/2}\exp\left\{-\tfrac{1}{2}\boldsymbol{\mu}'\Sigma^{-1}\left[I_p-(I_p-2\Sigma At)^{-1}\right]\boldsymbol{\mu}\right\}\\&\quad\times\phi_p\left[\mathbf{y};(I_p-2A\Sigma t)^{-1}\boldsymbol{\mu},\Sigma(I_p-2A\Sigma t)^{-1}\right],\end{aligned}$$

hence

$$\begin{aligned}Ee^{\mathbf{y}'A\mathbf{y}}&=|I_p-2A\Sigma t|^{-1/2}\exp\left\{-\tfrac{1}{2}\boldsymbol{\mu}'\Sigma^{-1}\left[I_p-(I_p-2\Sigma At)^{-1}\right]\boldsymbol{\mu}\right\}\\&\quad\times C\int\phi_p\left[\mathbf{y};(I_p-2A\Sigma t)^{-1}\boldsymbol{\mu},\Sigma(I_p-2A\Sigma t)^{-1}\right]\Phi_q\left[D(\mathbf{y}-\boldsymbol{\mu});\boldsymbol{\nu},\Delta\right]d\mathbf{y}\\&=C|I_p-2A\Sigma t|^{-1/2}\exp\left\{-\tfrac{1}{2}\boldsymbol{\mu}'\Sigma^{-1}\left[I_p-(I_p-2\Sigma At)^{-1}\right]\boldsymbol{\mu}\right\}\\&\quad\times E\Phi_q\left[D(Y-\boldsymbol{\mu});\boldsymbol{\nu},\Delta\right],\end{aligned}$$

where $Y\sim N_p\left[(I_p-2A\Sigma t)^{-1}\boldsymbol{\mu},\Sigma(I_p-2A\Sigma t)^{-1}\right]$, thus the result follows applying Lemma 15 with $m=n=1$. ■

Proof of Proposition 11. The result follows by equation (6) and Theorem 1.2.22 of Gupta and Nagar (2000). ■

Proof of Corollary 12. From equations (1.2.3) and (1.2.5) of Gupta and Nagar (2000) we get that $K_{mp}^{-1} = K_{pm}$ and $K_{pm}(\Psi \otimes \Sigma) K_{mp} = \Sigma \otimes \Psi$. The assertion follows from Proposition 11. ■

Proof of Corollary 13. From Theorem 1.2.21 (iv) and equations (1.2.3) and (1.2.5) of Gupta and Nagar (2000) we get that $(A \otimes B)' = A' \otimes B'$, $K_{qn}^{-1} = K_{nq}$ and $K_{pn}(I_n \otimes D') K_{qn} = D' \otimes I_n$, thus $K_{pn}(I_n \otimes D') = K_{pn}(I_n \otimes D') K_{qn} K_{nq} = (D' \otimes I_n) K_{nq}$. What it lacks of the proof is continued from Proposition 11. ■

Proof of Proposition 14. From Theorem 1.2.22 of Gupta and Nagar (2000) we get that $\operatorname{vec}(W') = (A_1 \otimes A_2') \operatorname{vec}(Y')$, the result follows from (6).

Proof of Lemma 15. Let $U \sim N_{q,n}(L, Q)$ a random matrix independent of W. Then

$$\begin{aligned}
E_W[\Phi_{q,n}(A + BWC; L, Q)] &= E_W \Pr(U \leq A + BWC | W) \\
&= \Pr(U - BWC \leq A) \\
&= \Phi_{q,n}(A; L - BMC, Q + (B \otimes C') S (B' \otimes C)) \\
&= \Phi_{q,n}(A + BMC; L, Q + (B \otimes C') S (B' \otimes C))
\end{aligned}$$

The last part of the proof is due to

$$U - BWC \sim N_{q,n}[L - BMC; Q + (B \otimes C') S (B' \otimes C)].$$

■

For related results see Zacks (1981), pp. 53-54.

Proof of Proposition 16. From equation (1.2.6) of Gupta and Nagar (2000) we get that $\operatorname{tr}(Y'T) = (\operatorname{vec}(T'))' \operatorname{vec}(Y')$ and due to $\operatorname{vec}(Y') \sim CSN_{pm;qn}(\operatorname{vec}(M'), \Sigma \otimes \Psi, D \otimes E, \operatorname{vec}(L'), Q)$ the rest of the proof follows from Lemma 1 of González-Farías *et al* (2004b).

Proof of Proposition 17. From equation (1.2.6) of Gupta and Nagar (2000) we get that $\operatorname{tr}(AY'BYCT) = (\operatorname{vec}(Y'))' (B \otimes (CT'A)) \operatorname{vec}(Y')$ the result follows from Proposition 5.

Proof of Corollary 18. Using the specified values of the parameters of the distribution of Y in (13) we get

$$\begin{aligned}
M_Z(T) = &\frac{\Phi_1[0, 0, \vartheta + (D \otimes E) \Theta (D' \otimes E')]}{\Phi_1[0, 0, \vartheta + (D\Sigma D') \otimes (E\Psi E')]} \\
&\times |I_{pm} - 2I_p \otimes (\Psi T')|^{-1/2},
\end{aligned}$$

which simplifies to

$$M_Z(T) = |I_q - 2\Psi T'|^{-p/2}.$$

■

Proof of Corollary 19.. Given that T is a real number we deduce that $\Theta = [I_q - 2(\Sigma^{-1}\Sigma) \otimes (T'\Psi)]^{-1} = [I_q - 2I_q \otimes T]^{-1} = [I_q - 2I_q T]^{-1} = (1-2T)^{-1} I_q$. Now

$$\begin{aligned} M_Z(T) &= \frac{\Phi_{qn}\left[0,0,Q+(D\otimes E)(1-2T\Psi)^{-1} I_q (D'\otimes E')\right]}{\Phi_{qn}[0,0,Q+(D\Sigma D')\otimes(EE')]} \\ &\times \left|I - 2\left(\Sigma\Sigma^{-1}\right)\otimes T\right|^{-1/2} \\ &= \frac{\Phi_{qn}\left[0,0,Q+(1-2T)^{-1}((DD')\otimes(EE'))\right]}{\Phi_{qn}[0,0,Q+(D\Sigma D')\otimes(EE')]} \\ &\times |I_q - 2I_q T|^{-1/2} \\ &= (1-2T)^{-q/2}. \end{aligned}$$

■

BIBLIOGRAPHY

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

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

Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, **83**, 715-726.

Copas, J.B. and Li, H.G. (1997). Inference for Non-random Samples. *Journal of the Royal Statistical Society,* B, 59, 1, 55-95.

Díaz-García, J.A. and González-Farías (2005) Singular matrix variate skew-elliptical distribution and the distribution of general linear transformation. *Comunicación Técnica No I-05-03/25-01-2005 (PE/CIMAT)*

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.

Domínguez-Molina J.A., González-Farías G., and Ramos-Quiroga R. (2004). Skew Normality in Stochastic Frontier Analysis. (Genton M. editor) Edited volume on "Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality". 223-241.

Genton, M. (2004), *Skew-elliptical Distributions and Their Applications: A Journey Beyond Normality*, (Editor), Chapman and Hall/CRC.

González-Farías, G., Domínguez-Molina, J. A. and Gupta, A.K. (2004a). The Closed Skew Normal Distribution. In: Genton M. G. ed., *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality.* Chapman and Hall/CRC, 25-42.

González-Farías, G., Domínguez-Molina, J. A. and Gupta, A. K. (2004b). Additive properties of skew normal random vectors. *Journal of Statistical Planning and Inference.* 126 (2004) 521-534.

Gupta, A.K. and Nagar, D.K. (2000). Matrix Variate Distributions. Chapman & Hall/CRC.

Gupta, A.K., Gonzáles-Farías, G. and Domínguez-Molina, J.A. (2003). A multivariate skew normal distribution. *Journal of Multivariate Analysis.* 89 (2004) 181-190.

Horn, R.A. and Johnson, C.R. (1985). *Matrix Analysis.* Cambridge University Press.

Kumbhakar, S.C. and Lovell, C.A. (2000). *Stochastic Frontier Analysis.* Cambridge: Cambridge University Press.

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.

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

Zacks, S. (1981). *Parametric Statistical Inference.* Oxford, Pergamon.