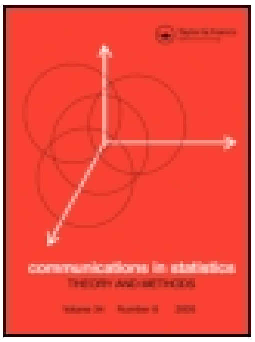

Taylor & Francis
Taylor & Francis Group

Communications in Statistics—Theory and Methods

ISSN: 0361-0926 (Print) 1532-415X (Online) Journal homepage: http://www.tandfonline.com/loi/lsta20

A Matrix Variate Closed Skew-Normal Distribution with Applications to Stochastic Frontier Analysis

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

To cite this article: J. Armando Domínguez-Molina , Graciela González-Farías , Rogelio Ramos-Quiroga & Arjun K. Gupta (2007) A Matrix Variate Closed Skew-Normal Distribution with Applications to Stochastic Frontier Analysis, Communications in Statistics—Theory and Methods, 36:9, 1691-1703, DOI: 10.1080/03610920601126126

To link to this article: https://doi.org/10.1080/03610920601126126

Published online: 30 Jun 2007.

Communications in Statistics—Theory and Methods, 36: 1691–1703, 2007

ISSN: 0361-0926 print/1532-415X online
DOI: 10.1080/03610920601126126

Distribution Theory

A Matrix Variate Closed Skew-Normal Distribution with Applications to Stochastic Frontier Analysis

J. ARMANDO DOMÍNGUEZ-MOLINA[1],
GRACIELA GONZÁLEZ-FARÍAS[2],
ROGELIO RAMOS-QUIROGA[2],
AND ARJUN K. GUPTA[3]

[1]Universidad de Guanajuato, Guanajuato, Mexico
[2]Centro de Investigación en Matemáticas, Guanajuato, Mexico
[3]Bowling Green State University, Bowling Green, Ohio, USA

In this article, we introduce the matrix extension of the closed skew-normal distribution and give two constructions for it: a marginal one and another based on hidden truncation. Important basic properties of the distribution are presented such as its closure under linear transformation and moment generating function. We also give distributional results for quadratic forms involving random matrices distributed according to two particular cases of it. Using an additive construction, we derive a submodel which can be employed to describe the compound error structure of a very general multivariate stochastic frontier model. Finally, we consider the skew-elliptical extension of the proposed distribution.

Keywords Chi-square distribution; Compound error; Hidden truncation; Linear transformation; Quadratic forms; Skew-elliptical distribution; Wishart distribution.

Mathematics Subject Classification Primary 62E15; Secondary 62H10.

1. Introduction

Various multivariate skew-normal distributions have been proposed in the literature, with each one of them aiming to characterize a particular aspect of a given phenomenon. For example, one emphasizes invariance under quadratic forms, another one uses a general latent structure to define distributions, etc.; see Genton (2004) for an overview. Nevertheless, most of these skew-normal distributions are special cases of the closed skew-normal (CSN) family of distributions as defined in Domínguez-Molina et al. (2003). The CSN class of distributions is closed under

Received July 31, 2005; Accepted September 28, 2006
Address correspondence to J. Armando Domínguez-Molina, Universidad de Guanajuato, Jalisco s/n, Guanajuato, 36240 Mexico; E-mail: jadguez@cimat.mx

the operations of marginalization and conditioning basic to statistical modeling, includes the normal distribution, and enjoys some of the appealing properties of the latter. In particular, the expressions for its marginal and conditional densities are similar to those for the normal case. However, the distributions included in the CSN class are, in general, skewed.

Here we consider the extension of the CSN distribution from the vector to the matrix case. The distribution we propose implicitly defines the matrix variate generalizations of many other multivariate skew-normal in the literature, and permits the inclusion of dependence structures, such as those for panel data, which are basic to the analysis of stochastic frontier models.

The articles by Aigner et al. (1977) and Meeusen and van Den Broeck (1977) were seminal to the development of models capable of describing the production efficiency of companies. In them, the concept of a stochastic frontier was introduced via the model $y = f(\mathbf{x}; \boldsymbol{\beta}) + \varepsilon$, where the error term, $\varepsilon = v - u$, is composed of a symmetric disturbance term, v, which represents measurement error, and by the non-negative, firm-specific term u which captures technical inefficiencies. This formulation of the error structure seeks to explain how companies with the same technical ability to manage their resources might end up with different output levels, due to the unobservable shocks v. Developments over the last 30 years in the specification and estimation of frontier production functions are discussed in Coelli et al. (2005).

Assuming a cross-sectional data structure, Domínguez-Molina et al. (2004) proposed a stochastic frontier model based on the CSN distribution as given in González-Farías et al. (2004a). Their proposal encompasses nested submodels with an increasing degree of complexity for the covariance structure, but within the framework of normal measurement errors and truncated normals for inefficiencies. Specifically, their model is

$$\mathbf{y} = \mathbf{f}(X; \boldsymbol{\beta}) + \mathbf{v} + G\mathbf{u}, \tag{1}$$

where $\mathbf{y}$ is a vector consisting of the value-added values for p firms, $\mathbf{f}$ is the production function commonly based on the Cobb-Douglas model with lagged input variables, $\mathbf{v} \sim N_p(\mathbf{0}, \Sigma)$ models measurement error, and $\mathbf{u} \sim N_q^{\mathbf{c}}(\boldsymbol{\nu}, \Lambda)$, $q \geq p$, where $N_p^{\mathbf{c}}(\boldsymbol{\nu}, \Lambda)$ denotes the $N_q(\boldsymbol{\nu}, \Lambda)$ distribution truncated below at $\mathbf{c}$. The random vector $\mathbf{u}$ models technological inefficiencies in groups of firms, and is weighted by the $p \times q$ full row rank matrix G. Also, it is assumed that $\mathbf{v}$ is independent of $\mathbf{u}$, $\mathbf{f}(X; \boldsymbol{\beta}) = (f(\mathbf{x}_1; \boldsymbol{\beta}), \ldots, f(\mathbf{x}_p; \boldsymbol{\beta}))'$, $X = (\mathbf{x}_1, \ldots, \mathbf{x}_p)'$ 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 $G = I_p$ or $G = -I_p$ for firm-specific cost efficiencies or technical inefficiencies, respectively.

For reference purposes, we repeat the formal definition of the density of the CSN distribution given originally by Domínguez-Molina et al. (2003).

Definition 1.1. 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

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

with

$$C^{-1} = \Phi_q(\mathbf{0}; \boldsymbol{\nu}, \Delta + D\Sigma D'), \tag{2}$$

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, respectively. In the definition of these last two functions, $\boldsymbol{\eta} \in \mathbb{R}^p$ denotes a mean vector and Ψ a $p \times p$ covariance matrix.

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

Domínguez-Molina et al. (2004) show that the density of the compound error term in model (1), $\boldsymbol{\varepsilon} = \mathbf{v} + G\mathbf{u}$, is $g(\boldsymbol{\varepsilon}) = \Phi_q^{-1}(\mathbf{0}; \mathbf{c} - \boldsymbol{\nu}, \Lambda)\phi_p(\boldsymbol{\varepsilon}; G\boldsymbol{\nu}, \Theta)\Phi_q[\Lambda G'\Theta^{-1}(\boldsymbol{\varepsilon} - G\boldsymbol{\nu}); \mathbf{c} - \boldsymbol{\nu}, \Upsilon]$, where $\Theta = \Sigma + G\Lambda G'$ and $\Upsilon = \Lambda - \Lambda G'\Theta^{-1}G\Lambda$. Thus,

$$\boldsymbol{\varepsilon} \sim CSN_{p,q}(G\boldsymbol{\nu}, \Theta, \Lambda G'\Theta^{-1}, \mathbf{c} - \boldsymbol{\nu}, \Upsilon).$$

The data structure for $\mathbf{y}_t = (y_{1t}, y_{2t}, \ldots, y_{pt})'$, $t = 1, \ldots, m$, in (1) is assumed to be that of a cross-sectional sample of p firms. For panel data structures of the form $Y_{p\times m} = (\mathbf{y}_1, \ldots, \mathbf{y}_m)$, where the p firms are followed through times $t = 1, \ldots, m$, we propose, in Sec. 4 of the article, a multivariate stochastic frontier model that, in principle, can capture general correlation patterns through time and between firms. More specifically, we consider the model $Y = F + \Xi$, where F is a matrix production function and $\Xi = V + DUE'$ is a matrix variate compound error structure. With the latter structure in mind, we first develop the matrix variate closed skew-normal distribution and study its basic properties. Thus, the remainder of the article is organized as follows. In Sec. 2, we give technical results for the CSN distribution that can be directly applied in stochastic frontier analysis (SFA). These results are also required for the matrix variate representation. Section 3 provides the definition and basic properties of the matrix variate closed skew-normal distribution, and presents results for quadratic forms involving random matrices of this type. In Sec. 4, we employ the matrix variate skew-normal distribution in the definition of our highly flexible model for use in SFA. In Sec. 5, we provide some concluding remarks and indicate some directions for further research. The proofs for two of the propositions presented in the article are given in the Appendix.

2. The Closed Skew-Normal Distribution and Its Properties

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 do the sums of independent CSN random variables. These closure properties allow one to study the distributional properties of random samples in a tractable way, and are very useful when considering the extension to the matrix variate case under certain types of dependencies. In what follows, we give various results which, apart from being of interest in themselves, also provide the building blocks for the matrix variate extension and the investigation of its properties.

The moment generating function of the CSN distribution, given in González-Farías et al. (2004a), allows us to easily derive the moments of the distribution and to prove important distributional results. It is given in closed form as

$$M_{\mathbf{y}}(\mathbf{s}) = \frac{\Phi_q(D\Sigma\mathbf{s}; \boldsymbol{\nu}, \Delta + D\Sigma D')}{\Phi_q(\mathbf{0}; \boldsymbol{\nu}; \Delta + D\Sigma D')} e^{\mathbf{s}'\boldsymbol{\mu} + \frac{1}{2}\mathbf{s}'\Sigma\mathbf{s}}, \quad \mathbf{s} \in \mathbb{R}^p. \tag{3}$$

The following proposition gives an alternative marginal representation of the CSN distribution which is useful, for instance, when conducting simulation or calculating moments. Moreover, the probabilistic structure defined within it can be applied directly in stochastic frontier modeling. A simpler version of this result was given in Domínguez-Molina et al. (2004).

Proposition 2.1 (Marginal Representation). *Let* $\mathbf{v} \sim N_p(\mathbf{0}, I_p)$, $\mathbf{u} \sim N_q(\mathbf{0}, \Delta + D\Sigma D')$ *and* $\mathbf{u}$ *be independent of* $\mathbf{v}$. *Then the distribution of*

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

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

This representation in terms of normals and truncated normals is far more general than other representations given in the literature in terms of sums. Moreover, it proves to be very flexible when modeling different error structures for the SFA model.

An alternative way of motivating the closed skew-normal distribution is via a hidden truncation process which, in many applications, will be highly plausible. For example, when the observational mechanism for measuring a variable is such that we only record a value when an external condition is satisfied, an asymmetric distribution will often be induced. The hidden truncation characterization also furnishes a useful means of establishing some of the properties of skew distributions, by so doing providing greater insight as to how they arise. For the hidden truncation process, we first condition a normal random vector on a set of latent variables subject to certain given restrictions (e.g., $Z \geq 0$), thus generating a CSN distribution. Then, if we consider operations such as marginalization, conditioning, or addition, their application results in distributions which are also members of the CSN family. However, it is important to point out that we can reverse this procedure in the following way. First, carry out the corresponding marginalization, conditioning, or addition procedure on the normal random vector and then consider the hidden truncation process. This will lead to exactly the same distribution, as shown in Domínguez-Molina et al. (2003). The same argument applies when we obtain the joint distribution of independent CSN random variables.

Using the conditioning approach of Domínguez-Molina et al. (2003), we provide a simple derivation of the distribution function of a CSN random vector which proves to be useful in the study of dependence structures via copulas (Nelsen, 2006).

Proposition 2.2. *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 as given in* (2).

Although the conditioning argument provides a means with which to derive elegant proofs for certain results, it cannot be used, for instance, in the calculation of moments. For the latter, the representation in terms of sums is far more useful. Hence, we will use the marginal representation given in Proposition 2.1 when considering the application of the matrix variate extension of the CSN distribution to SFA.

3. The Matrix Variate CSN Distribution and its Properties

In this section we introduce the matrix variate generalization of the CSN distribution.

First, we define the $p \times m$ random matrix of observations as

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

where $\mathbf{x}_i(p \times 1)$, $i = 1, \dots, m$ is the ith column of X. Here, $\mathbf{x}_1, \dots, \mathbf{x}_m$ can be thought as a sample of size m from a p-dimensional population, but it is not necessary to assume that $\mathbf{x}_1, \dots, \mathbf{x}_m$ are independent. Now, for a $p \times m$ matrix X, the vec operator, $\mathrm{vec}(X)$, is the $mp \times 1$ vector defined as

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

The random matrix X is said to have a matrix variate normal distribution with mean matrix $M(p \times m)$ and covariance matrix $\Omega(mp \times mp)$ if $\mathrm{vec}(X') \sim N_{pm}(\mathrm{vec}(M'), \Omega)$. We will use the notation $X \sim N_{p,m}(M, \Omega)$ and denote the probability density function (pdf) and the cumulative distribution function (cdf) of X as $\phi_{p,m}(X; M, \Omega) = \phi_{pm}(\mathrm{vec}(X); \mathrm{vec}(M), \Omega)$, and $\Phi_{p,m}(X; M, \Omega) = \Phi_{pm}(\mathrm{vec}(X); \mathrm{vec}(M), \Omega)$.

Using the above and the material on the CSN distribution presented in the preceding two sections, we are now in the position to define its matrix variate extension.

Definition 3.1. A random matrix $Y(p \times m)$ is said to have a matrix variate closed skew-normal (MVCSN) distribution with parameters $M(p \times m)$, $S(mp \times mp)$, $B(nq \times mp)$, $L(q \times n)$, and $Q(nq \times nq)$, with $S > 0$ and $Q > 0$, if

$$\mathrm{vec}(Y') \sim CSN_{pm,qn}[\mathrm{vec}(M'), S, B, \mathrm{vec}(L'), Q].$$

We will use the notation

$$Y \sim CSN_{p,m;q,n}(M, S, B, L, Q) \tag{4}$$

to denote the fact. In most cases, the matrices S and B will have specific structures. Properties for the parametrization (4) are obtained immediately from González-Farías et al. (2004a).

When working with random matrices it is important to bear in mind how the random matrix, Y, is assembled. Here we consider the situation in which $Y = (\mathbf{y}_1, \ldots, \mathbf{y}_n)$ is a sample of independent and identically distributed (iid) $CSN_{p,q}(\boldsymbol{\mu}, \Sigma, D, \boldsymbol{v}, \Delta)$ random vectors. Due to Corollary 2.4.1 of González-Farías et al. (2004b), we know that the distribution of $\mathrm{vec}(Y) = (\mathbf{y}'_1, \ldots, \mathbf{y}'_n)'$ is

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

and hence

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

Thus, assuming iid columns for Y we obtain the distribution of Y', not that of Y as we might have hoped for. In order to obtain the distribution of Y, we first consider the distribution of the transpose of a MVCSN matrix. We start by defining the commutation matrix which transforms $\mathrm{vec}(A)$ into $\mathrm{vec}(A')$. The commutation matrix, $K_{mp}(mp \times mp)$, is defined as $K_{mp} = \sum_{i=1}^{m} \sum_{j=1}^{p} (H_{ij} \otimes H'_{ij})$, where the (i, j)th element of $H_{ij}(m \times p)$ is 1 and all its other elements are 0. Then, if

$$X \sim CSN_{p,m;q,n}(M, S, B, L, Q),$$

the distribution of X' can be obtained from the fact that $\mathrm{vec}(X) = K_{mp}\mathrm{vec}(X')$. Using Theorem 1 of González-Farías et al. (2004b) and Theorem 1.2.22 of Gupta and Nagar (2000), we then obtain that

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

Moreover, if $S = \Sigma \otimes \Psi$, with $\Sigma(p \times p) > 0$ and $\Psi(m \times m) > 0$, then

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

This follows because, from Eqs. (1.2.3) and (1.2.5) of Gupta and Nagar (2000), $K_{mp}^{-1} = K_{pm}$ and $K_{pm}(\Psi \otimes \Sigma)K_{mp} = \Sigma \otimes \Psi$. Finally, returning to the distribution of $Y = (\mathbf{y}_1, \ldots, \mathbf{y}_n)$, we can use the above results to obtain

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

Alternatively, a matrix variate CSN distribution can be obtained using an extension of the hidden truncation argument of Copas and Li (1997). This construction, which may be more natural in many experimental settings, proceeds as follows.

Define the independent normal random matrices $U_1 \sim N_{p,m}(0, S)$ and $U_2 \sim N_{q,n}(0, Q)$, where, as previously, S is $mp \times mp$ and Q is $nq \times nq$. Now, consider the matrices $W = M + U_1$ and $Z = -L + DU_1E' + U_2$, where D is $q \times p$, E is $n \times m$ and, as before, M is $p \times m$ and L is $q \times n$. Then the joint distribution of W and Z is

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

where

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

Now, if $Y \overset{d}{=} W \mid \{Z \geq 0\}$, we obtain that $f(Y) = K\phi_{p,m}(Y; M, S)\Phi_{q,n}[E(Y-M)D'; L, Q]$, where $K^{-1} = \Phi_{q,n}[0; L, Q + (D \otimes E)S(D' \otimes E')]$. Hence,

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

which is a particular case of (4).

3.1. *Basic Properties*

Here we present certain basic properties of the MVCSN distribution. First we consider the distribution of linear transformations of MVCSN matrices and then give the distribution's moment generating function. The section ends with results for quadratic forms of MVCSN variables.

3.1.1. *Linear Transformation.* Here we consider a closure property for linear transformations of MVCSN matrices of the form $W = A_1YA_2$. This kind of transformation admits contrasts among rows as well as among columns which, for the usual setting of random matrices, would allow contrasts among individuals and among attributes.

Proposition 3.1. *Consider $Y \sim CSN_{p,m;q,n}(M, S, B, L, Q)$ and let $A_1(n_1 \times p)$ and $A_2(m \times n_2)$ be matrices such that $A = A_1 \otimes A_2'$ has full row rank. If $W = A_1YA_2$ then*

$$W \sim CSN_{n_1,n_2;q,n}(M_A, S_A, B_A, L, Q_A),$$

where $M_A = A_1MA_2$, $S_A = ASA'$, $B_A = BSA'S_A^{-1}$, and $Q_A = Q + BSB' - BSA'S_A^{-1}ASB'$.

Proof. Using Theorem 1.2.22 of Gupta and Nagar (2000) we obtain that $\text{vec}(W') = (A_1 \otimes A_2')\text{vec}(Y')$. The result then follows from Theorem 1 of González-Farías et al. (2004b).

3.1.2. *Moment Generating Function.* Prior to presenting the moment generating function (mgf) of the MVCSN distribution, we need to introduce some additional notation. We consider the partitioned matrices $B = (B_1', \ldots, B_q')'$ and $S = (S_1', \ldots, S_m')'$, where B_i is $n \times mp$, $i = 1, \ldots, q$ and S_j is $p \times mp$, $j = 1, \ldots, m$. Let $T(p \times m)$ be an arbitrary matrix, $T^\dagger = [B_1S\,\text{vec}(T'), \ldots, B_qS\,\text{vec}(T')]$ and $S^\dagger = [S_1\,\text{vec}(T'), \ldots, S_m\,\text{vec}(T')]$.

Proposition 3.2. *Let $Y \sim CSN_{p,m;q,n}(M, S, B, L, Q)$. Then the mgf of Y is given by*

$$M_Y(T) = E\,\text{etr}(Y'T) = \frac{\Phi_{q,n}(T^\dagger; L', Q + BSB')}{\Phi_{q,n}(0; L', Q + BSB')}\text{etr}\left(M'T + \frac{1}{2}S^{\dagger\prime}T\right) \tag{5}$$

where, for A a square matrix, $\text{etr}(A) = \exp[\text{trace}(A)]$.

Proof. Due to the fact that $\mathrm{tr}(Y'T) = (\mathrm{vec}(T'))'\mathrm{vec}(Y')$, and also that $\mathrm{vec}(Y') \sim CSN_{pm;qn}(\mathrm{vec}(M'), S, B, \mathrm{vec}(L'), Q)$, we obtain from (3) that

$$E\,\mathrm{etr}(Y'T) = \frac{\Phi_{nq}(BS\,\mathrm{vec}(T'); \mathrm{vec}(L'), Q + BSB')}{\Phi_{nq}(\mathbf{0}; \mathrm{vec}(L'), Q + BSB')} \times \exp\left[(\mathrm{vec}(T'))'\mathrm{vec}(M') + \frac{1}{2}(\mathrm{vec}(T'))'S\,\mathrm{vec}(T')\right]. \quad (6)$$

Now, by noting that $BS\mathrm{vec}(T') = \mathrm{vec}(T^{\dagger})$ and $S\,\mathrm{vec}(T') = \mathrm{vec}(S^{\dagger})$, we obtain $(\mathrm{vec}(T'))'S\,\mathrm{vec}(T') = \mathrm{tr}(S^{\dagger\prime}T)$. Finally, (5) results by making use of these results, together with the definition of $\Phi_{q,n}(\cdot)$ given at the beginning of Sec. 3, in (6). □

The mgf for the MVCSN distribution with the parametrization $S = \Sigma \otimes \Psi$ and $B = D \otimes E$, where $\Sigma(p \times p)$ and $\Psi(m \times m)$ are positive definite and $D(n \times p)$ and $E(q \times m)$ are arbitrary matrices, is given by the following corollary.

Corollary 3.1. *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*

$$M_Y(T) = \frac{\Phi_{q,n}(E\Psi T'\Sigma D'; L, Q + (D\Sigma D') \otimes (E\Psi E'))}{\Phi_{q,n}[0; L, Q + (D\Sigma D') \otimes (E\Psi E')]}\mathrm{etr}\left(M'T + \frac{1}{2}T'\Sigma T\Psi\right).$$

3.1.3. *Quadratic Forms.* As is well known, the distributional properties of quadratic forms of normal variables play a key role in classical inference. Certain results for quadratic forms of skew-normal variates have appeared recently in the literature. Azzalini and Capitanio (1999, Sec. 3.3), discuss the independence of quadratic forms and present a theorem which is similar to the Fisher-Cochran Theorem given in Rao (1973, Sec. 3b.4). Loperfido (2001) considers quadratic forms for skew-normal random vectors. Genton et al. (2001) derive the moments of skew-normal random vectors and their quadratic forms, and consider applications in time series analysis and spatial statistics. Finally, Wang et al. (2004) establish an equivalence between the chi-square and generalized skew-normal distributions. They also show how properties of the chi-square distribution extend to the univariate and multivariate skew-normal distributions. In what follows, we present three results related to the quadratic forms of MVCSN matrices. As will become evident, these results draw heavily on the work of Domínguez-Molina et al. (2003) on quadratic forms of CSN variates.

Proposition 3.3. *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

$$M_Z(T) = \frac{\Phi_{q,n}[0, L, Q + (D \otimes E)\Theta(D' \otimes E')]}{\Phi_{q,n}[0, L, Q + (D\Sigma D') \otimes (E\Psi E')]}|I_{mp} - 2(\Sigma B) \otimes (\Psi CT'A)|^{-1/2}, \quad (7)$$

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

Proof. From Eq. (1.2.6) of Gupta and Nagar (2000), we obtain that $\mathrm{tr}(AY'BYCT) = (\mathrm{vec}(Y'))'(B \otimes (CT'A))\mathrm{vec}(Y')$. The result then follows from Proposition 13 of Domínguez-Molina et al. (2003).

Corollary 3.2. *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 a Wishart distribution with parameters m, p, and Ψ, that is $Y'\Sigma^{-1}Y \sim W_m(p, \Psi)$.*

Proof. Using the specified values of the parameters of the distribution of Y in (7), we obtain that

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

which simplifies to $M_Z(T) = |I_m - 2\Psi T'|^{-p/2}$.

Note that, as a direct consequence of Corollary 3.2, if $\mathbf{y}$ has a $CSN_{p,1}(\mathbf{0}, \Sigma, \boldsymbol{\delta}, 0, 1)$ distribution then $\mathbf{y}\mathbf{y}' \sim W_p(1, \Sigma)$.

Corollary 3.3. *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 $Y'\Sigma^{-1}Y \sim \chi^2_p$.*

Proof. Given that T is a real number, we deduce that $\Theta = [I_p - 2(\Sigma^{-1}\Sigma) \otimes T']^{-1} = [I_p - 2I_p \otimes T]^{-1} = [I_p - 2I_pT]^{-1} = (1 - 2T)^{-1}I_p$. Now,

$$\begin{aligned} M_Z(T) &= \frac{\Phi_{np}[\mathbf{0}, \mathbf{0}, Q + (D \otimes E)(1 - 2T\Psi)^{-1}I_p(D' \otimes E')]}{\Phi_{np}[\mathbf{0}, \mathbf{0}, Q + (D\Sigma D') \otimes (EE')]}|I_p - 2(\Sigma\Sigma^{-1}) \otimes T|^{-1/2} \\ &= \frac{\Phi_{np}[\mathbf{0}, \mathbf{0}, Q + (1 - 2T)^{-1}((DD') \otimes (EE'))]}{\Phi_{np}[\mathbf{0}, \mathbf{0}, Q + (D\Sigma D') \otimes (EE')]}|I_p - 2I_pT|^{-1/2} \\ &= (1 - 2T)^{-p/2}. \end{aligned}$$

4. A Multivariate Stochastic Frontier Model

In this section we extend the relationship between the closed skew-normal and SFA models to the matrix case using a similar approach to that used by Domínguez-Molina et al. (2004) for the vector case.

In what follows, we will use the notation $U \sim N^C_{m,n}(M, S)$ to denote that U is a $N_{m,n}(M, S)$ random matrix truncated below at C. That is, the truncation is of the type $U \geq C$, where $W \geq C$ means $W_{ij} \geq C_{ij}$, $i = 1, \ldots, m$, $j = 1, \ldots, n$. Note that $U \geq C \Rightarrow \mathrm{vec}(U') \geq \mathrm{vec}(C')$.

Consider, now, production data on p firms at time t. We 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 $\mathbf{f}(X_t; \boldsymbol{\beta}_t) = (f(\mathbf{x}_{1t}; \boldsymbol{\beta}_t), \ldots, f(\mathbf{x}_{pt}; \boldsymbol{\beta}_t))'$, $X_t = (\mathbf{x}_{1t}, \ldots, \mathbf{x}_{pt})'$ is a known matrix of covariates, $\boldsymbol{\beta}_t$ is unknown, $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \ldots, \varepsilon_{pt})'$ is a random vector of compound errors and $\boldsymbol{\varepsilon}_t = \mathbf{v}_t + G\mathbf{u}_t$, with $\mathbf{v}_t = (v_{1t}, \ldots, v_{pt})'$, $\mathbf{u}_t = (u_{1t}, \ldots, u_{qt})'$ and G is a $p \times q$ weighting matrix. We use Y to denote the $p \times m$ matrix of the value added observations for the p

firms at times $t = 1, \ldots, m$, i.e.,

$$Y = \begin{pmatrix} y_{11} & \cdots & y_{1m} \\ \vdots & \ddots & \vdots \\ y_{p1} & \cdots & y_{pm} \end{pmatrix} = (\mathbf{y}_1, \ldots, \mathbf{y}_m).$$

A joint model for such production data can be written as

$$Y = F + \Xi, \tag{8}$$

where $F = (\mathbf{f}(X_1, \boldsymbol{\beta}_1), \ldots, \mathbf{f}(X_m, \boldsymbol{\beta}_m))$, $\Xi = V + GU$, $V = (\mathbf{v}_1, \ldots, \mathbf{v}_m)$, and $U = (\mathbf{u}_1, \ldots, \mathbf{u}_m)$.

We choose, in fact, to consider a slightly more general model for the compound errors, namely,

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

where $V \sim N_{p,m}(0, S)$, $U \sim N^C_{q,m}(L, Q)$, $D(p \times q)$, $E(m \times m)$, and V is independent of U. By pre-multiplying the matrix of technical inefficiencies, U, by D we can incorporate common inefficiencies within groups of similar companies. Similarly, by post-multiplying U by E', time related inefficiency effects can be allowed for. Note that the matrix V is no longer constrained to merely reflect measurement error. Indeed, depending on the structure of the variance matrix S, it can also incorporate random effects such as random intercepts and time-induced correlations among the columns of Y. Given that $\mathrm{vec}(\Xi') = \mathrm{vec}(V') + (D \otimes E)\mathrm{vec}(U')$, we obtain from Domínguez-Molina et al. (2004) that the density of the compound error $\Xi = V + DUE'$ is

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

where $\Theta = S + (D \otimes E)Q(D' \otimes E')$ and $\Upsilon = Q - Q(D' \otimes E')\Theta^{-1}(D \otimes E)Q$. Thus, Ξ has a matrix variate closed skew-normal distribution. Specifically,

$$\Xi \sim CSN_{p,m;q,m}(DLE', \Theta, Q(D' \otimes E')\Theta^{-1}, C - L, \Upsilon).$$

Model (8), with the compound error structure $\Xi = V + DUE'$, includes the following submodels as special cases:

- Model I (Homoscedastic and uncorrelated errors). D an arbitrary $p \times q$ matrix, $E = I_m$, $S = I_m \otimes \Sigma$, and $Q = I_m \otimes \Delta$, where $\Sigma(p \times p)$ and $\Delta(q \times q)$ are covariance matrices.
- Model II (Heteroscedastic and uncorrelated errors). D an arbitrary $p \times q$ matrix, $E = I_m$, S and Q block diagonal matrices of the form $S = \bigoplus_{i=1}^m \Sigma_i$, and $Q = \bigoplus_{i=1}^m \Delta_i$, with $\bigoplus$ denoting the matrix direct sum operator (see Horn and Johnson, 1985, p. 24). The result of $A \oplus B$ is a block diagonal matrix. Here, $\Sigma_i(p \times p)$ and $\Delta_i(m \times m)$ are covariance matrices, $i = 1, \ldots, m$.
- Model III (Correlated errors). If any of the matrices E, S, or Q are non block diagonal.

5. Concluding Remarks and Directions of Future Research

In this article, we have emphasized the close relationship that exists bewteen stochastic frontier analysis and skew distributions, in particular the CSN distribution and its matrix extension the MVCSN distribution. The latter permits flexible structures such as those for panel data with contemporaneous as well as temporal dependencies.

In our discussion of the distributional properties of the CSN and MVCSN distributions we considered quadratic forms. One line of potential future research would be to establish conditions for the existence of independent quadratic forms that would generalize the results of Propositions 8 and 9 of Azzalini and Capitanio (1999).

A classical approach to estimating the parameters of the MVCSN distribution would be to use the maximum likelihood method. However, even for the univariate and multivariate cases of the skew-normal distribution, a range of problems are known to exist when using this approach; see Pewsey (2000) and Azzalini (2005). For the MVCSN distribution, the problems associated with maximum likelihood estimation are likely to be far more complex due to the elevated number of parameters requiring estimation. Hence, the reliable estimation of its parameters represents a challenging problem for future research.

An appealing direct extension of the MVCSN distribution would be to allow for the inclusion of important distributional features such as heavy tails. This could be done by incorporating the skew-elliptical family of distributions, as we now describe.

A random matrix $Y(p \times m)$ is said to have a matrix variate extended skew-elliptical (MVESE) distribution with pdf generator h and parameters $M(p \times m)$, $S(mp \times mp)$, $B(nq \times mp)$, $L(q \times n)$, $Q(nq \times nq)$, where $S > 0$ and $Q > 0$, if

$$\mathrm{vec}(Y') \sim ESE_{pm,nq}[\mathrm{vec}(M'), S, B, \mathrm{vec}(L'), Q, h].$$

Here, ESE denotes the extended skew-elliptical distribution as given in González-Farías et al. (2004a). We denote this relation by

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

Using similar arguments to those employed in Sec. 3, it is also possible to derive the matrix variate skew-elliptical distribution for which the parameter matrix $B = D \otimes E$, where $D(n \times p)$ and $E(q \times m)$ are arbitrary matrices.

Appendix

Proof of Proposition 2.1. In order to obtain the distribution of $\mathbf{y}$ we use the mgf technique. Now,

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

Using the Sherman–Morrison–Woodbury formula, we obtain $(\Sigma^{-1} + D'\Delta^{-1}D)^{-1} + \Sigma D'(\Delta + D\Sigma D')^{-1}D\Sigma = \Sigma$. Thus,

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

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

Proof of Proposition 2.2. By definition, $F_{p,q}(\mathbf{y}_0; \boldsymbol{\mu}, \Sigma, D, \boldsymbol{\nu}, \Delta) = \Pr(\mathbf{y} \leq \mathbf{y}_0)$. Now, from the extension of the Copas and Li model given in González-Farías et al. (2004a), we obtain 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 on noting that

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

□

Acknowledgments

This work was partially supported by the research projects 45974-F CONACYT-México and SEJ2004-04101-ECON (Spanish Ministry of Education) and by CONCYTEG grant 04-16-K117-028.

References

Aigner, D., Lovell, C. A. K., Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *J. Econometrics* 6:21–37.

Azzalini, A. (2005). The skew-normal distribution and related multivariate families. *Scand. J. Statist.* 32:159–188.

Azzalini, A., Capitanio, A. (1999). Statistical applications of the multivariate skew-normal distributions. *J. Roy. Statist. Soc. B* 61:579–602.

Coelli, T. J., Prasada Rao, D. S., O'Donnell, C. J., Battese, G. (2005). *An Introduction to Efficiency and Productivity Analysis*. New York: Springer Science+Business Media.

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

Domínguez-Molina, J. A., González-Farías, G., Gupta, A. K. (2003). The multivariate closed skew normal distribution. Technical Report No. 03-12, Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH.

Domínguez-Molina, J. A., González-Farías, G., Ramos-Quiroga, R. (2004). Skew normality in stochastic frontier analysis. In: Genton, M. G., ed. *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*. Boca Raton, FL: Chapman and Hall/CRC, pp. 223–241.

Genton, M. G., ed. (2004). *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*. Boca Raton, FL: Chapman and Hall/CRC.

Genton, M. G., He, L., Liu, X. (2001). Moments of skew-normal random vectors and their quadratic forms. *Statist. Probab. Lett.* 51:319–325.

González-Farías, G., Domínguez-Molina, J. A., Gupta, A. K. (2004a). The closed skew-normal distribution. In: Genton M. G., ed. *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*. Boca Raton, FL: Chapman and Hall/CRC, pp. 25–42.

González-Farías, G., Domínguez-Molina, J. A., Gupta, A. K. (2004b). Additive properties of skew normal random vectors. *J. Statist. Plann. Infer.* 126:521–534.

Gupta, A. K., Nagar, D. K. (2000). *Matrix Variate Distributions*. Boca Raton, FL: Chapman & Hall/CRC.

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

Loperfido, N. (2001). Quadratic forms of skew-normal random vectors. *Statist. Probab. Lett.* 54:381–387.

Meeusen, W., van Den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *Int. Economic Rev*. 18:435–444.

Nelsen, R. B. (2006). *An Introduction to Copulas*. 2nd ed. New York: Springer Science+Business Media.

Pewsey, A. (2000). Problems of inference for Azzalini's skew-normal distribution. *J. Appl. Statist*. 27:859–870.

Rao, C. R. (1973). *Linear Statistical Inference and its Applications*. 2nd ed. New York: Wiley.

Wang, J., Boyer, J., Genton, M. G. (2004). A note on the equivalence between chi-square and generalized skew-normal distributions. *Statist. Probab. Lett.* 66:395–398.