7-2-2004

On Ranking and Selection from Independent Truncated Normal Distributions

William C. Horrace
Syracuse University, whorrace@maxwell.syr.edu

On Ranking and Selection from Independent Truncated Normal Distributions

William Horrace, Syracuse University

Keywords: Ranking and selection; Truncated normal; Stochastic frontier

Abstract

This paper develops probability statements and ranking and selection rules for independent truncated normal populations. An application to a broad class of parametric stochastic frontier models is considered, where interest centers on making probability statements concerning unobserved firm-level technical inefficiency. In particular, probabilistic decision rules allow subsets of firms to be deemed relatively efficient or inefficient at prespecified probabilities. An empirical example is provided. r 2004 Elsevier B.V. All rights reserved.

JEL classification: C12; C16; C44; D24

1. Introduction

Truncated distributions are common in economics, where non-negative random variables characterize data generation processes. On the most fundamental level, price and quantity are assumed to be non-negative. A specific distributional class, truncated normal distributions, are commonly used. For example, truncated normal are used in the censored and truncated regression models, see Tobin (1958), Amemiya (1974) or Heckman (1976). Recently, Hong and Shum (2002) show that the vector of drop-out prices observed in an asymmetric ascending auction may be multivariate truncated normal. Additionally, the truncated normal distribution is used to describe technical inefficiency in parametric stochastic frontier models, see Greene (2005). The importance (and relative complexity) of multivariate truncated normals is illustrated in the sizeable Bayesian literature devoted to computer simulation of these random variates, for example, see Geweke (1991).

This paper develops probability statements on independent truncated normal distributions that characterize the relative magnitude of realizations from the distributions. That is, if $\mathbf{W}' = [W_1, \ldots, W_n]$ is a multivariate truncated normal random variable, the goal is to attach probabilities to statements on the relative magnitudes of the Wj when they are assumed independent. In particular, this paper presents probabilistic ranking and selection rules to determine subsets of the n elements of W that are relatively small or large at prespecified probabilities. The proposed rules are based on a non-standard multivariate distribution derived from differences of independent truncated normals. While the form of the multivariate distribution is non-standard (complicated by the truncation), the probability inequalities are readily calculable. An application to parametric stochastic frontiers models is considered; these models yield truncated normal distributions for firm-level technical (in)efficiency, and then attempt to characterize the ranks of realizations from these (in)efficiency distributions. The proposed selection rules accomplish this task by identifying relatively (in)efficient firms at a prespecified probability, and it is argued that the rules are more theoretically justified than current methods for assessing distributional differences in these models.

2. Characterizations of the distribution

This section presents characterizations of the multivariate truncated normal distribution.

Definition 1. Let $\mathbf{W}^* = [W_1^*, \ldots, W_n^*]'$ be an n-dimensional random variable. W^* has a non-singular n-variate normal distribution with mean vector $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_n]'$ and (n x n) positive definite covariance matrix $\boldsymbol{\Sigma}$, if it has density

$$f_{\mathbf{W}^*}(\mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-n/2}(\det \boldsymbol{\Sigma})^{-1/2} \exp[-\tfrac{1}{2}(\mathbf{w} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{w} - \boldsymbol{\mu})], \quad \mathbf{w} \in \mathbb{R}^n.$$

Adopt the standard notation: W_ Nð0;RÞ. Let $\mathbf{W}^* \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})$. Let $\mathbf{W} = [W_1, \ldots, W_n]'$ be thetruncation of W^* below $\mathbf{c} = [c_1, \ldots, c_n]'$, $\mathbf{c} \in \mathbb{R}^n$.

Definition 2. W has an n-variate truncated normal distribution given by

$$f_{\mathbf{W}}(\mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{(2\pi)^{-n/2}(\det \boldsymbol{\Sigma})^{-1/2} \exp\{-\frac{1}{2}(\mathbf{w}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{w}-\boldsymbol{\mu})\}}{(2\pi)^{-n/2}(\det \boldsymbol{\Sigma})^{-1/2} \int_{\mathbf{c}}^{\infty} \exp\{-\frac{1}{2}(\mathbf{w}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{w}-\boldsymbol{\mu})\}\,d\mathbf{w}}, \quad \mathbf{w} \in \mathbb{R}^n_{\geqslant \mathbf{c}}$$
$$= \frac{\exp\{-\frac{1}{2}(\mathbf{w}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{w}-\boldsymbol{\mu})\}}{\int_{\mathbf{c}}^{\infty} \exp\{-\frac{1}{2}(\mathbf{w}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{w}-\boldsymbol{\mu})\}\,d\mathbf{w}}, \quad \mathbf{w} \in \mathbb{R}^n_{\geqslant \mathbf{c}},$$

Where $\int_{\mathbf{a}}^{\mathbf{b}}$ is an n-dimensional Riemann integral from a $\mathbf{a}\ (n \times 1)$to $\mathbf{b}\ (n \times 1)$and$\mathbb{R}^n_{\geqslant \mathbf{c}} = \{\mathbf{w} \in \mathbb{R}^n : \mathbf{w} \geqslant \mathbf{c}\}$. One could envision truncation of a subset of $\mathbf{W}^*$ this just requires that for certain W_j^*, the truncation point goes from c_j to $-\infty$ in the limit. Other forms of truncation have been suggested by Tallis (1963, 1965) and Beattie (1962). Define $(n \times 1)$ vectors: $\mathbf{M} = \mathbf{c} - \boldsymbol{\mu}$, $\mathbf{t} \in \mathbb{R}^n$ and $\mathbf{P}\,(n \times 1) = \mathbf{c} - \boldsymbol{\mu} - \imath\boldsymbol{\Sigma}\mathbf{t}$ with typical elements $M_j,\ t_j,\ P_j,\ j \in N,\ N = \{1, \ldots, n\}$ respectively, and $\imath = \sqrt{-1}$.. Then, the characteristic function of W is given by the following result:

Theorem 3. The characteristic function of W is

$$CF_{\mathbf{W}}(\mathbf{t}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{\int_{\mathbf{M}}^{\infty} \exp\{-\frac{1}{2}(\mathbf{w} - \imath\boldsymbol{\Sigma}\mathbf{t})'\boldsymbol{\Sigma}^{-1}(\mathbf{w} - \imath\boldsymbol{\Sigma}\mathbf{t})\}\,d\mathbf{w}}{\int_{\mathbf{M}}^{\infty} \exp\{-\frac{1}{2}\mathbf{w}'\boldsymbol{\Sigma}^{-1}\mathbf{w}\}\,d\mathbf{w}} \exp\{\imath\mathbf{t}'\boldsymbol{\mu} - \tfrac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}\}$$
$$= \frac{\int_{\mathbf{P}}^{\infty} \exp\{-\frac{1}{2}\mathbf{u}'\boldsymbol{\Sigma}^{-1}\mathbf{u}\}\,d\mathbf{u}}{\int_{\mathbf{M}}^{\infty} \exp\{-\frac{1}{2}\mathbf{u}'\boldsymbol{\Sigma}^{-1}\mathbf{u}\}\,d\mathbf{u}} \exp\{\imath\mathbf{t}'\boldsymbol{\mu} - \tfrac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}\}.$$

The proof is in Section A.1. Tallis (1961) derives a similar formula for the moment generating function of W. The derivation hinges on a series of variable transformations that shift W without scaling it. The univariate case $(n = 1)$ was first suggested in a problem posed by Horrace and Hernandez (2001). Of course, the characteristic function generates the moments of W. Tallis (1961) derives these moments using differentiation of the moment generating function when $\mathbf{W}^* \sim \mathrm{N}(\mathbf{0}, \mathbf{R})$ and R is the correlation matrix associated with $\boldsymbol{\Sigma}$. Amemiya (1974) adapts the Tallis results for the case where $\mathbf{W}^* \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma})$. Weiler (1959) derives them for the case n = 2 using integration of the density f_W. While most of the results of this paper can be derived for arbitrary truncation below c, this paper is concerned with truncation below 0 for each element of W^*. Moreover, ranking and selection rules are greatly simplified by orthogonality. Therefore, we will always make the following two assumptions.

Assumption A.0. c = 0.

Assumption A.1. $\boldsymbol{\Sigma}$ is a diagonal matrix with typical element σ_j^2

.

Definition 4. Under A.0 and A.1 the elements of W* are independent, and W has an n-variate truncated normal distribution given by

$$f_{\mathbf{W}}(\mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{(2\pi)^{-n/2}\sigma^{-1} \exp\{-\frac{1}{2}\sum_{j=1}^{n} \frac{(w_j - \mu_j)^2}{\sigma_j^2}\}}{\prod_{j=1}^{n}[1 - \Phi(-\mu_j/\sigma_j)]}, \quad w_j \in \mathbb{R}_+,$$

where

$$\sigma^2 = \prod_{j=1}^{n} \sigma_j^2 \text{ and } \mathbb{R}^n_+ = \{\mathbf{w} \in \mathbb{R}^n : \mathbf{w} \geqslant \mathbf{0}\}$$

This leads to a useful formula for the cumulative distribution function.

Definition 5. Under A.0 and A.1 the cumulative distribution function (c.d.f.) of W is

$$F_{\mathbf{W}}(\mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{\prod_{j=1}^{n}[\Phi(\{w_j - \mu_j\}/\sigma_j) - \Phi(-\mu_j/\sigma_j)]}{\prod_{j=1}^{n}[1 - \Phi(-\mu_j/\sigma_j)]}, \quad w_j \in \mathbb{R}_+.$$

The assumptions also lead to a similar simplification of the characteristic function.

Corollary 6. Under A.0 and A.1, the characteristic function of W is

$$CF_{\mathbf{W}}(\mathbf{t}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{\prod_{j=1}^{n}[1 - \Phi(-\mu_j/\sigma_j - \imath\sigma_j t_j)]}{\prod_{j=1}^{n}[1 - \Phi(-\mu_j/\sigma_j)]} \exp\left\{\sum_{j=1}^{n}\left(\imath t_j\mu_j - \frac{1}{2}t_j^2\sigma_j^2\right)\right\}$$

It is clear from Definition 4 or Corollary 6 that:

Conclusion 7. Wj and Wi are independent if and only if W_j^* and W_i^*

The result can be generalized to the case where Wj and Wi are uncorrelated vectors of truncated normal variates which are individually correlated within themselves. This is identical to the result for multivariate normal distributions, see Tong (1990, Theorem 3.3.2). Under A.0 and A.1 the first two moments of the distribution are

$$\mathrm{E}(W_j) = \xi_j = \mu_j + \frac{\sigma_j\phi(-\mu_j/\sigma_j)}{1 - \Phi(-\mu_j/\sigma_j)} \tag{1}$$

And

$$Var(W_j) = \sigma_j^2\left\{1 - \frac{\mu_j}{\sigma_j}\left[\frac{\phi(-\mu_j/\sigma_j)}{[1 - \Phi(-\mu_j/\sigma_j)]}\right] - \left[\frac{\phi(-\mu_j/\sigma_j)}{[1 - \Phi(-\mu_j/\sigma_j)]}\right]^2\right\}. \tag{2}$$

These are widely known results; for example, see Bera and Sharma (1999). A useful monotonicity result is:

Lemma 8. *Under* A.1:

(i) $\mathrm{E}(W_j) = \xi_j$ *is increasing in* μ_j *and*
(ii) $\mathrm{E}(W_j) = \xi_j$ *is increasing in* σ_j.

The proof is in Section A.1. Result (i) of Lemma 8 was indirectly shown by Bera and Sharma (1999). The implication is that if $\mu_j > \mu_I$ for fixed $\sigma_j = \sigma$ $\sigma_j > \sigma_i$ for fixed $\mu_j = \mu$, then $\xi_j > \xi_i$ for $j \neq i$. Therefore, in certain cases, the relative ranks of ξ_j can be assessed by examining the relative ranks of μ_j or σ_j. This is potentially useful, if the distribution of some estimates of the μ_j are normal or asymptotically normal. If so, ranking inference on the estimates of μ_j would be standard, while ranking inference on the ξ_j, using the transformation of the estimates of μ_j in Eq. (1), would be non-standard.

3. Linear transformations

The selection procedures that follow hinge on distributions of differences of truncated normals, so understanding the effects of linear transformations on the truncated normal distribution is useful. First, the family of multivariate truncated normal distributions (Definition 2) is not closed to linear transformations in general. Rescaling and/or summation of elements cause the distribution to loose its truncated normal shape. However, it is closed to relocation. Second, even under Assumption A.1, the family of multivariate truncated normal distributions (Definition 4) is not closed to linear transformations. However, under A.1 it is closed to relocation and (positive) rescaling. (Positive rescaling is only necessary to preserve truncation below the truncation point; negative rescaling produces truncation above the truncation point.) Consequently, the sampling distribution for the sample average

from a random sample of a truncated normal population will not be truncated normal.[1] Third, the marginal distributions from multivariate truncated normal distributions will not be truncated normal in general, however under the independence assumption (A.1) the marginal distributions are truncated normal.1 The consequence of the preceding is that ranking and selection rules for differences of independent truncated normals will hinge on non-standard distributions and, in particular, not truncated normal distributions. For example, under A.0 and A.1, the density of the difference $W_j - W_k,\ j \neq k$ is

$$
\begin{aligned}
f_{W_j - W_k}(w) &= \begin{cases} \int_0^\infty f_{W_j}(x, \mu_j, \sigma_j^2) f_{W_k}(x - w, \mu_k, \sigma_k^2)\,\mathrm{d}x & \text{for } w < 0 \\ \int_0^\infty f_{W_j}(w + x, \mu_j, \sigma_j^2) f_{W_k}(x, \mu_k, \sigma_k^2)\,\mathrm{d}x & \text{for } w \geqslant 0 \end{cases} \\
&= \begin{cases} \dfrac{1 - \Phi((\sigma_j^2 + \sigma_k^2)^{-1/2}[-\frac{\sigma_j}{\sigma_k}(w + \mu_k) - \mu_j \frac{\sigma_k}{\sigma_j}])}{\{2\pi\sigma_j^2\sigma_k^2\}^{1/2}[1 - \Phi(-\mu_j/\sigma_j)][1 - \Phi(-\mu_k/\sigma_k)]} & \\ \qquad \times \exp\left\{ - \dfrac{[w - (\mu_j - \mu_k)]^2}{2(\sigma_j^2 + \sigma_k^2)} \right\} & \text{for } w < 0 \\ \dfrac{1 - \Phi((\sigma_j^2 + \sigma_k^2)^{-1/2}[\frac{\sigma_k}{\sigma_j}(w - \mu_j) - \mu_k \frac{\sigma_j}{\sigma_k}])}{\{2\pi\sigma_j^2\sigma_k^2\}^{1/2}[1 - \Phi(-\mu_j/\sigma_j)][1 - \Phi(-\mu_k/\sigma_k)]} & \\ \qquad \times \exp\left\{ - \dfrac{[w - (\mu_j - \mu_k)]^2}{2(\sigma_j^2 + \sigma_k^2)} \right\} & \text{for } w \geqslant 0, \end{cases}
\end{aligned}
\tag{3}
$$

where f Wj is the marginal density function for Wj given in Definition 4 with n = 1. The partition of the integral on w<0 and w⩾0 is for computational convenience.[2] Notice that $W_j - W_k \in \mathbb{R}$, so the density is not truncated at all. Also, when $\mu_j = \mu_k$ and $\sigma_j = \sigma_k$, the distribution is symmetric about the origin. Consider generalizing Eq. (3) to the (n-1)-dimensional case where *k* is a control index. Define $\boldsymbol{\delta}^k \in \mathbb{R}^{n-1}$ vector

$$
\boldsymbol{\delta}^k = [W_1 - W_k, \ldots, W_{k-1} - W_k, W_{k+1} - W_k, \ldots, W_n - W_k]'.
$$

Let $\mathbf{d} = [d_1, \ldots, d_{n-1}]'$ be any realization of $\boldsymbol{\delta}^k$, then under A.0 and A.1 the distribution of $\boldsymbol{\delta}^k$ is

$$
\begin{aligned}
F_{\boldsymbol{\delta}^k}(\mathbf{d}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \Pr\{\boldsymbol{\delta}^k \leqslant \mathbf{d}\} \\
&= \int_{\mathbf{w}:\boldsymbol{\delta}^k \leqslant \mathbf{d}} f_{\mathbf{W}}(\mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\Sigma})\,\mathrm{d}\mathbf{w} \\
&= \int_{\max_{j \neq k}(-d_j, 0)}^{\infty} \prod_{j \neq k}^{n} F_{W_j}(w_k + d_j, \mu_j, \sigma_j^2) f_{W_k}(w_k, \mu_k, \sigma_k^2)\,\mathrm{d}w_k,
\end{aligned}
\tag{4}
$$

where FWj is the marginal distribution function for Wj given by Definition 5 with n = 1. The upper tail probabilities are

---

[1] Proofs of all of the preceding facts are in Horrace (2003), available from the author.

[2] If we start with wj and wk, and then transform to w = wj - wk and x = wj , we must have x ⩾ ⩾0 and x ⩾ w, so we handle both by integrating over x from zero to infinity, when w<0. Conversely, if we let x = wk we need x ⩾0 and x ⩾ - w and so this is the more convenient transformation when w ⩾0.

$$
\begin{aligned}
F^*_{\delta^k}(\mathbf{d}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \Pr\{\delta^k \geqslant \mathbf{d}\} \\
&= \int_{\mathbf{W}:\delta^k \geqslant \mathbf{d}} f_{\mathbf{W}}(\mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\Sigma})\, \mathrm{d}\mathbf{w} \\
&= \int_0^\infty \int_{\max_{j\neq k}\{w_k + d_j, 0\}}^\infty \prod_{i=1}^{n} f_{W_i}(w_i, \mu_i, \sigma_i^2)\, \mathrm{d}\mathbf{w} \\
&= \int_0^\infty \prod_{j\neq k}^{n} \left[1 - F_{W_j}\left(\max_{j\neq k}\{w_k + d_j, 0\}, \mu_j, \sigma_j^2\right)\right] \\
&\quad \times f_{W_k}(w_k, \mu_k, \sigma_k^2)\, \mathrm{d}w_k. \qquad (5)
\end{aligned}
$$

The probabilities given in Eqs. (4) and (5) are general (can be used for any independent, absolutely continuous distribution with no probability mass below zero) and are used in the next section to derive the selection rules. The equations can be used to construct multivariate probability statements on the $W_j - W_k,\ j \neq k$. Define a $\alpha \in (0, 1)$ and L_k^α as the solution in dj = d (for all j) to $F^*_{\delta^k}(\mathbf{d}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = 1 - \alpha$. Also, define $\gamma \in (0, 1)$ and U_k^γ as the solution in dj = d (for all j) to $F_{\delta^k}(\mathbf{d}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \gamma$ Then, L_k^α and U_k^γk are one-sided confidence bounds similar to the multiple comparisons with a control (MCC) intervals suggested by Dunnett (1955), although here there is no sample of which to speak. Dunnett made inferential statements for the sampling distribution of population statistics; these statements are for individual realizations from the underlying truncated normal populations. Notice that in general $F_{\delta^k}(\mathbf{d}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \neq 1 - F^*_{\delta^k}(\mathbf{d}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$.

4. Selection rules

Suppose that we are interested in the relative ranks of a potential realization from the distribution of W under A.0 and A.1. Let

$$w_{[n]} > w_{[n-1]} > \cdots > w_{[1]}$$

be the ranks of the elements of a single potential realization of W. Interest centers on selecting a subset of the indices {1; 2; . . . ; n} that contains the index [n] with a prespecified confidence level and another subset that contains [1] with a prespecified confidence level. Consider the following selection rules Rmax and Rmin:

$$
\begin{aligned}
R_{\max}: &\quad \text{Select population } k \text{ if } U_k^{1-\gamma} \leqslant 0, \\
R_{\min}: &\quad \text{Select population } k \text{ if } L_k^{\alpha} \geqslant 0.
\end{aligned}
$$

Furthermore, define corresponding subsets Smax and Smin:

$$
\begin{aligned}
S_{\max} &= \{k : U_k^{1-\gamma} \leqslant 0\}, \\
S_{\min} &= \{k : L_k^{\alpha} \geqslant 0\}.
\end{aligned}
$$

Notice that monotonicity of $F^*_{\delta^k}$ and F_{δ^k}k in d implies equivalent selection rules

$$
\begin{aligned}
R_{\max}: &\quad \text{Select population } k \text{ if } 1 - \gamma \leqslant F_{\delta^k}(\mathbf{0}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \Pr\{[n] = k\}, \\
R_{\min}: &\quad \text{Select population } k \text{ if } 1 - \alpha \leqslant F^*_{\delta^k}(\mathbf{0}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \Pr\{[1] = k\}. \qquad (6)
\end{aligned}
$$

The continuity of δ^k ensures that

$$\Pr\{([1]=k)\cap([1]=j),\ j\neq k\}=\Pr\{([n]=k)\cap([n]=j),\ j\neq k\}=0.$$

That is, there can only be one minimum or maximum with positive probability. Therefore, probability statements such as

$$\begin{aligned}\Pr\{[n]\in S_{\max}|R_{\max}\}&=\sum_{i\in S_{\max}}\Pr\{[n]=i\}=\sum_{i\in S_{\max}}F_{\delta^i}(\mathbf{0},\boldsymbol{\mu},\boldsymbol{\Sigma}),\\ \Pr\{[1]\in S_{\min}|R_{\min}\}&=\sum_{i\in S_{\min}}\Pr\{[1]=i\}=\sum_{i\in S_{\min}}F^*_{\delta^i}(\mathbf{0},\boldsymbol{\mu},\boldsymbol{\Sigma})\end{aligned}\qquad(7)$$

are valid. Of course, when Smax is empty, the probability given in Eq. (7) is just 0; similarly, for Smin. Let us always assume:

$$\text{A.2: } \gamma<0.5 \text{ and } \alpha<0.5.$$

The following is a useful result:

Lemma 9. *S_{max} can have no more that one element. Similarly, Smin can have no more that one element.*

Proof. *Suppose not*. If there were more that one index in Smax, then there would be more than one index that satisfies Rmax, so there would be more than one index k where $F_{\delta^k}(\mathbf{0},\boldsymbol{\mu},\boldsymbol{\Sigma})\geqslant 1-\gamma$. Therefore, $\Pr\{[n]\in S_{\max}|R_{\max}\}>1$ under A.2 in Eq. (7). *Contradiction*. The proof is completed similarly for Smin.

Of course Smax and Smin can be empty, so there are only two states for the subsets: empty set or singleton. Given Lemma 9, Eq. (7) becomes

$$\begin{aligned}\Pr\{[n]\in S_{\max}|R_{\max}\}&=F_{\delta^{(n)}}(\mathbf{0},\boldsymbol{\mu},\boldsymbol{\Sigma}),\\ \Pr\{[1]\in S_{\min}|R_{\min}\}&=F^*_{\delta^{(1)}}(\mathbf{0},\boldsymbol{\mu},\boldsymbol{\Sigma}),\end{aligned}\qquad(8)$$

where (n) and (1) correspond to the indices contained in Smax and Smin, respectively, when the subsets are both singletons. That is, Smax = {(n)} and Smin {(1)}. This leads to the following result:

Theorem 10. *Let $F_{\delta^{(n)}}(\mathbf{0},\boldsymbol{\mu},\boldsymbol{\Sigma})$ and $F^*_{\delta^{(1)}}(\mathbf{0},\boldsymbol{\mu},\boldsymbol{\Sigma})=1-\alpha_{(1)}$ and $F^*_{\delta^{(1)}}(\mathbf{0},\boldsymbol{\mu},\boldsymbol{\Sigma})=1-\alpha_{(1)}$. If Smax in nonempty, then the probability of a correct selection conditional on the selection rule Rmax is*

$$\Pr\{[n]\in S_{\max}|R_{\max}\}=1-\gamma_{(n)}\geqslant 1-\gamma.$$

Similarly, if Smin is non-empty, then the probability of a correct selection conditional on the selection rule Rmin is

$$\Pr\{[1]\in S_{\min}|R_{\min}\}=1-\alpha_{(1)}\geqslant 1-\alpha.$$

Otherwise, there is no inference at the prespecified level.

Proof. The proof follows directly from Eqs. (6) and (8).

For a prespecified confidence level $(1-\gamma \text{ or } 1-\alpha)$, a correct selection is guaranteed at that level (as long as the inference is defined). Theorem 10 is related to the results of Gupta (1965), but Gupta's results are based on a sample of observations. Rizvi (1971) considers ranking and selection statements for the absolute value of estimates of μ_j, but the result of Theorem 10 is markedly different[3]

The subsets S_max and S_min contain those single indices with high probability of corresponding to the maximal Wj and minimal Wj , respectively. One could consider

[3] Rizvi considers ex post truncation (the absolute value of an estimate of μ_j that is normal); here the truncation is ex ante, so normality (and its invariance to relocation and rescaling) is not available to simplify the probability statements.

finding subsets containing indices with low probability of corresponding to the maximal Wj and minimal Wj . Therefore, alternative (but not equivalent) rules are

$$R_{-\max}: \quad \text{Select population } k \text{ if } U_k^{\gamma} \geqslant 0,$$
$$R_{-\min}: \quad \text{Select population } k \text{ if } L_k^{1-\alpha} \leqslant 0,$$

with corresponding subsets S_max and S_min. (Note the notational subtlety: ‘‘max’’ corresponds to ‘‘maximum with high probability’’, while ‘‘ max’’ corresponds to ‘‘maximum with low probability’’ or ‘‘not the maximum’’.) Again, monotonicity implies

$$R_{-\max}: \quad \text{Select population } k \text{ if } \gamma \geqslant F_{\delta^k}(\mathbf{0}, \boldsymbol{\mu}, \boldsymbol{\Sigma}),$$
$$R_{-\min}: \quad \text{Select population } k \text{ if } \alpha \geqslant F^*_{\delta^k}(\mathbf{0}, \boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

A useful result that relates selection rules and subsets is:

Lemma 11. The sets Smax and S_max are non-intersecting. Also, the sets Smin and S_min are *non-intersecting*.

Proof. Under A.2, monotonicity of F_{δ^k} implies $1-\gamma > \gamma$ so that $U_k^{1-\gamma} \geqslant U_k^{\gamma}$ Therefore, if Rmax selects k $(0 \geqslant U_k^{1-\gamma})$, then R_max will not select *k*, because $0 \geqslant U_k^{1-\gamma} \geqslant U_k^{\gamma}$ violates the selection rule: $U_k^{\gamma} \geqslant 0$ (and vice versa). The proof is completed similarly for Smin and S_min.

Here we define the probabilities of correct selection as: $\Pr\{[n] \notin S_{-\max} | R_{-\max}\}$and $\Pr\{[1] \notin S_{-\min} | R_{-\min}\}$, respectively. Then

$$\Pr\{[n] \notin S_{-\max} | R_{-\max}\} = \sum_{i \notin S_{\max}} \Pr\{[n] = i\} = \sum_{i \notin S_{\max}} F_{\delta^i}(\mathbf{0}, \boldsymbol{\mu}, \boldsymbol{\Sigma}),$$
$$\Pr\{[1] \notin S_{-\min} | R_{-\min}\} = \sum_{i \notin S_{\min}} \Pr\{[1] = i\} = \sum_{i \notin S_{\min}} F^*_{\delta^i}(\mathbf{0}, \boldsymbol{\mu}, \boldsymbol{\Sigma}). \qquad (9)$$

These probabilities are not necessarily bound by the prespecified confidence level, unless Smax and S_max are non-empty.

Theorem 12. *If Smax is non-empty, the probability of a correct selection conditional on the selection rule* $R_{-\max}$ *is* $\Pr\{[n] \notin S_{-\max} | R_{-\max}\} \geqslant 1 - \gamma_{(n)} \geqslant 1 - \gamma$. *If Smin is nonempty, the probability of a correct selection conditional on the selection rule R min is* $\Pr\{[1] \notin S_{-\min} | R_{-\min}\} \geqslant 1 - \alpha_{(1)} \geqslant 1 - \alpha$.

Proof. When Smax is non-empty,

$$\Pr\{[n] \notin S_{-\max} | R_{-\max}\} \geqslant \Pr\{[n] \in S_{\max} | R_{\max}\} = 1 - \gamma_{(n)} \geqslant 1 - \gamma$$

because $S_{\max} \cap S_{-\max} = \emptyset$ Lemma 11. Therefore,

$$\Pr\{[n] \notin S_{-\max} | R_{-\max}\} \geqslant \Pr\{[n] \in S_{\max} | R_{\max}\} = 1 - \gamma_{(n)} \geqslant 1 - \gamma,$$

by Theorem 10. The proof is completed similarly for S_min.

Of course, the event $\{[n] \notin S_{-\max} | R_{-\max}\}$ does not necessarily imply the event $\{[n] \in S_{\max} | R_{\max}\}$, so $\Pr\{[n] \notin S_{-\max} | R_{-\max}\}$ may not be exactly $1 - \gamma_{(n)}$. In fact, the exact value is governed by Eq. (9). When Smax is empty, the probability of a correct selection is not bound from below and is determined by Eq. (9), but this does not preclude a reasonable probability of correct selection. Examples are provided below:

Example 13. Suppose that under A.0 and A.1, n = 3; μ_1 =1; μ_2 = 2; μ_3 = 3; σ_1^2 $\sigma_2^2 = \sigma_3^2 = 1$. Consider $F_{\delta^k}(\mathbf{0}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $F^*_{\delta^k}(\mathbf{0}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$:

$$
\begin{aligned}
F_{\delta^1}(\mathbf{0}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \Pr\{[n] = 1\} = 0.055, \\
F_{\delta^2}(\mathbf{0}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \Pr\{[n] = 2\} = 0.225, \\
F_{\delta^3}(\mathbf{0}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \Pr\{[n] = 3\} = 0.719, \\
F^*_{\delta^1}(\mathbf{0}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \Pr\{[1] = 1\} = 0.696, \\
F^*_{\delta^1}(\mathbf{0}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \Pr\{[1] = 2\} = 0.248, \\
F^*_{\delta^1}(\mathbf{0}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \Pr\{[1] = 3\} = 0.056. \ :
\end{aligned}
$$

For $\gamma = 0.05,\ S_{\max} = S_{-\max} = S_{\min} = S_{-\min} = \emptyset,$ so none of the variables have high or low probability of being the maximum or minimum. There is no inference at the prespecified level.

Example 14. Now suppose in the previous example that there is less variability and $\sigma_1^2 = \bar{\sigma}_2^2 = \sigma_3^2$ 0:50. Then

$$
\begin{aligned}
\Pr\{[n] = 1\} &= 0.014, \\
\Pr\{[n] = 2\} &= 0.155, \\
\Pr\{[n] = 3\} &= 0.831, \\
\Pr\{[1] = 1\} &= 0.820, \\
\Pr\{[1] = 2\} &= 0.166, \\
\Pr\{[1] = 3\} &= 0.014.
\end{aligned}
$$

Smax is still empty, so none of the variables have high probability of being the maximum. However, S_max = {1}, since $F_{\delta^1} \leqslant \gamma$. Therefore, one can conclude that index 1 corresponds to the maximum with low probability. By Eq. (9), the probability of a correct selection conditional on R_max is equal to 0:831 + 0:155 = 0:968, which happens to be greater than $1 - \gamma_{(n)}$ = 0:831 and 1 – y = 0:95. However, Theorem 12 is not governing this high confidence level, because Smax is empty. Instead, the high confidence level is strictly an artifact of these particular distributional assumptions. Also, Smin = $\emptyset$ and S_min = {3}.

Example 15. Suppose $n = 4,\ \mu_1 = 1,\ \mu_2 = 1,\ \mu_3 = 1,\ \mu_4 = 2,\ \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$ = 0:50 :

$$
\begin{aligned}
\Pr\{[n] = 1\} &= 0.112, \\
\Pr\{[n] = 2\} &= 0.112, \\
\Pr\{[n] = 3\} &= 0.112, \\
\Pr\{[n] = 4\} &= 0.663, \\
\Pr\{[1] = 1\} &= 0.320, \\
\Pr\{[1] = 2\} &= 0.320, \\
\Pr\{[1] = 3\} &= 0.320, \\
\Pr\{[1] = 4\} &= 0.041.
\end{aligned}
$$

For y = 0:05; Smax = $\emptyset$.;. However, S _max = {1, 2, 3}. By Eq. (9) the probability of a correct selection conditional on R_max is equal to 0.663 which is less than 1 - y. This example demonstrates that the probability bounding of Theorem 12 requires Smax not be empty. Also, Smin = $\emptyset$; and S_min = {4}.

Example 16. Suppose $n = 4,\ \mu_1 = 1,\ \mu_2 = 1,\ \mu_3 = 4,\ \mu_4 = 4,\ \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 =$ 0:15:

$$
\begin{aligned}
\Pr\{[n] = 1\} &= 0.000, \\
\Pr\{[n] = 2\} &= 0.000, \\
\Pr\{[n] = 3\} &= 0.500, \\
\Pr\{[n] = 4\} &= 0.500, \\
\Pr\{[1] = 1\} &= 0.500, \\
\Pr\{[1] = 2\} &= 0.500, \\
\Pr\{[1] = 3\} &= 0.000, \\
\Pr\{[1] = 4\} &= 0.000.
\end{aligned}
$$

For y = 0:05; Smax = ∅;. This illustrates the impact of Lemma 9. The distributions of 3 and 4 are equally probable of generating the largest observation, but they are not both in Smax. Now S_max = {1; 2}. Therefore, one can conclude that indices 1 and 2 correspond to the maximum with low probability. By Eq. (9), the probability of a correct selection conditional on R_max is 0:5 + 0:5 = 1:0. Also, Smin = ∅ and S_min = {3; 4}.

Example 17. Suppose $n = 4,\ \mu_1 = 1,\ \mu_2 = 2,\ \mu_3 = 3,\ \mu_4 = 4,\ \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = 0{:}15$:

$$\begin{aligned}
&\Pr\{[n] = 1\} = 0.000,\\
&\Pr\{[n] = 2\} = 0.000,\\
&\Pr\{[n] = 3\} = 0.034,\\
&\Pr\{[n] = 4\} = 0.966,\\
&\Pr\{[1] = 1\} = 0.966,\\
&\Pr\{[1] = 2\} = 0.034,\\
&\Pr\{[1] = 3\} = 0.000,\\
&\Pr\{[1] = 4\} = 0.000.
\end{aligned}$$

For g ¼ 0:05; Smax ¼ f4g, so index 4 corresponds to the maximum with high probability. The probability of a correct selection conditional on Rmax is equal to $1 - \gamma_{(n)} = 0.966 > 1 - \gamma$ which is consistent with Theorem 10. Additionally, S_max = {1; 2; 3} and the probability of a correct selection conditional on R_max is also $1 - \gamma_{(n)} = 0.966 > 1 - \gamma$ per Theorem 12. Also, Smin = {1} and S_min = {2; 3; 4}. It should be noted that all the probabilities in Eqs. (7)–(9) could be estimated by rejection sampling from univariate normal variates. Indeed, all the preceding examples were verified with rejection sampling simulations. However, rejection sampling can be impractical. In fact, the simulations used to verify the examples were only feasible because the probability of rejection was low (μ_j large and positive). When rejection sampling is not feasible, there is a growing body of literature devoted to efficient sampling from troublesome truncated normal distributions. For example, see Geweke (1991). However, even these techniques are subject to potential problems and criticisms. Therefore, if the integration in Eqs. (4) and (5) is calculable, then theoretical implementations of these procedures may be superior to simulation approaches.[4] Also, it is clear from the preceding examples, that the subsets are a convenient way of summarizing probabilities. Whether one wishes to report the subsets based on prespecified confidence levels or the actual probabilities (F_{δ^*} and $F^*_{\delta^*}$), is a matter of taste.

5. Stochastic frontiers

In the literature on productivity and efficiency measurement, a common parametric class of production (cost) function estimators imply conditional distributions for technical inefficiency that are independent normal random variables, truncated below zero. Consider the specification

$$y_{jt} = g(x_{jt}, \beta) + v_{jt} - u_{jt}, \quad j = 1, \ldots, n, \ \ t = 1, \ldots, T, \tag{10}$$

where y_{jt} is productive output of firm *j* in period *t*; *g* is a production function that maps a vector of productive inputs x_{jt} into output through the unknown parameter vector β. (The production function *g* typically satisfies some additional assumptions that are unimportant to the current discussion.) The $v_{jt} \in \mathbb{R}$ are random variables representing stochastic shocks to the production process. Let the distribution of v_{jt} be that of an i.i.d. zero-mean normal random variable with variance σ_v^2. Furthermore, let

$$u_{jt} = \gamma(t) u_j, \quad j = 1, \ldots, n, \ \ t = 1, \ldots, T,$$

where the $u_j \in \mathbb{R}_+$ are positive random variables representing technical inefficiencies, and $\gamma(t) > 0$ is some positive, continuous parameterization of *t*. In this formulation of the production function, smaller *u* corresponds to better

[4] There is also a Bayesian inference literature that has grown out of the stochastic frontier literature. These techniques either directly or indirectly provide inference on relative ranks using Bayesian sampling techniques and are a viable alternative to the results presented here. For example, see Fernandez et al. (2002), Tsionas (2002), Kim and Schmidt (2000), and Koop et al. (1997).

production (given inputs) and higher efficiency. Let the distribution of uj be the absolute value of an i.i.d. zeromean normal random variable with variance σ_u^2. Additionally, let the x_{jt}, v_{jt}, and u_{jt} be independent across *j* and across *t*. This parametric ''stochastic frontier'' model has been extensively studied in various forms originating with Aigner et al. (1977) who restrict the model to $\gamma(t) = 1$ and $g(x_{jt}, \beta) = x_j'\beta$. Jondrow et al. (1982) study a formulation similar to Aigner, Lovell, and Schmidt. Battese and Coelli (1988) consider the case where $\gamma(t) = 1$ and $g(x_{jt}, \beta) = x_{jt}'\beta$. jtb. Kumbhakar (1990) extends this to the case where $\gamma(t) = 1 - \exp(bt + ct^2)$. Battese and Coelli (1992) consider $\gamma(t) = \exp\{-\eta(t - T)\}$, while Cuesta (2000) considers $u_{jt} = \exp\{-\eta_j(t - T)\}u_j$. Finally, Greene (2005) relaxes the parametric form of technical efficiency and allows for heterogeneity in g. In all these formulations, the parametric assumptions on v_{jt} and u_{jt} imply that u_{jt} conditional on $\varepsilon_{jt} = v_{jt} - u_{jt}$ is a normal variable truncated below zero (this is also the case when u_{jt} is i.i.d. exponential). For example, in the Battese and Coelli (1988) formulation, the distribution of u_j conditional on ε_{jt} is the truncation below zero of an $N(\mu_{*j}, \sigma_*^2)$ variable where

$$\mu_{*j} = -\frac{\sigma_u^2 \bar{\varepsilon}_j}{\sigma_u^2 + \sigma_v^2/T} \tag{11}$$

and

$$\sigma_*^2 = \frac{\sigma_u^2 \sigma_v^2}{T\sigma_u^2 + \sigma_v^2}, \tag{12}$$

where $\bar{\varepsilon}_j = T^{-1}\sum_{t=1}^{T} \varepsilon_{jt}$. Additionally, in all the formulations neither realizations nor estimates of realizations of ujt are available; only estimates of the mean and variance of ujt conditional on ε_{jt}. Continuing the example, Battese and Coelli suggest maximum-likelihood estimates of μ_{*j} and σ_*^2 (although alternative consistent estimates, like GLS, exist) based on a point estimate of β. If it is assumed that the value of the estimate of β equals the true value of β, then the sampling variability in the estimates of μ_{*j} and σ_*^2 can be ignored and the conditional distribution of uj is independent truncated normal, and an estimate of firm-level technical efficiency is the mean of the conditional truncated distribution: $E(u_j|\varepsilon_{jt})$. Indeed, Battese and Coelli (1988, p. 391) state, ''[w]e obtain the conditional distribution of the firm effect [uj], given the value of the random variables, $[\varepsilon_{jt} = v_{jt} - u_j]$, $t = 1, 2, \ldots, T$. This assumes that the values of the parameter β are known''. More recent formulations, based on time-varying technical inefficiency, suggest $E(u_{jt}|\varepsilon_{jt})$ as an estimate of technical efficiency in period *t*. However, the common thread in all these parametric formulations is that the sampling variability in the estimates of μ_{*j} and σ_*^2 is ignored and the mean of the conditional distribution of u serves as a point estimate of technical inefficiency.[5]

5.1. Ranking the conditional means

Empirical implementations of these parametric models are too many to name here. However, they typically assume that β is known and include some sort of ranking of the conditional means, $E(u_{jt}|\varepsilon_{jt})$, over *j* in each period *t* as a proxy for the ranking of the unobserved random variable, ujt. For example, see Horrace and Schmidt (1996). Unfortunately, $E(u_{jt}|\varepsilon_{jt})$ is a misleading point estimate for ujt. While smaller $E(u_{jt}|\varepsilon_{jt})$ may suggest smaller ujt, it may not be the case that ujt is small in any particular sample, even if $E(u_{jt}|\varepsilon_{jt})$ is small. Therefore, using $E(u_{jt}|\varepsilon_{jt})$ as a point estimate of u_{jt} may have its limitations. Indeed, a firm j with $E(u_{it}|\varepsilon_{it}) \simeq 0$ may be operating with u_{jt} much greater than 0 in any sample, y_{jt}; x_{jt}.

That being said, Theorems 10 or 12 are a better way to draw inferences on technical inefficiency (instead of examining the rankings of the conditional means). That is, use Theorem 10 to define a set S_{min} that contains the j with the smallest (unobserved) u_{jt} with probability at least $1 - \alpha$. The idea is that these parametric stochastic frontier models only produce distributions for u_{jt}, not u_{jt} itself, and as such the conditional mean $E(u_{jt}|\varepsilon_{jt})$ can only characterize the distribution of u_{jt}, and not the probability of a realization of u_{jt} of specific magnitude. However, Theorem 10 may be used to characterize the magnitude of the u_{jt} in a probabilistic sense, and this is all that the data can allow. Ultimately, the traditional approach of ranking the conditional means and the new approach suggested here are similar in that both follow from the relative magnitudes of the means of the underlying normal

[5] This remains an unsolved problem in the stochastic frontier literature. However, as pointed out by a referee, Bayesian approaches do address this issue in a meaningful way.

distributions before truncation. However, the difference in the two approaches is embodied in the fact that the latter takes into account the variance of the underlying distribution. As such, using Theorem 10 to identify efficiency is theoretically more appealing.

5.2. Texas electrical utility application

We examine a formulation of the Eq. (10) with time-invariant technical inefficiency, although the selection rules could be applied in the time-varying case on a period-by-period basis. Consider the model of Horrace and Schmidt (1996):

$$y_{jt} = x_{it}'\beta + v_{jt} - u_j, \quad j = 1, \ldots, n.$$

Under the assumptions that $v_{jt} \sim \text{i.i.d. N}(0, \sigma_v^2)$; $u_j \sim |\text{i.i.d. N}(0, \sigma_u^2)|$; and x_{jt}, v_{tj}, u_j independent, generalized least squares (GLS) yields consistent estimates σ_v^2 v and σ_u^2 which imply the conditional distribution of u_j . Specifically, $\hat{\mu}_{*j}$ and $\hat{\sigma}_*^2$ are consistent for μ_{*j} and σ_*^2 in Eqs. (11) and (12), respectively. Then the usual point estimates of technical efficiency based on Battese and Coelli (1988) are

$$TE_j = \text{E}(\exp\{-u_j\}|\varepsilon_j) = \exp\left\{-\hat{\mu}_{*j} + \frac{1}{2}\hat{\sigma}_*^2\right\} \frac{1 - \Phi(\hat{\sigma}_* - \frac{\hat{\mu}_{*j}}{\hat{\sigma}_*})}{1 - \Phi(\frac{-\hat{\mu}_{*j}}{\hat{\sigma}_*})}. \tag{13}$$

Horrace and Schmidt (1996) calculate the GLS technical efficiency of 10 Texas electric utility plants from a panel of data between 1966 and 1985, where inputs to the production of the logarithm of electricity are capital, labor, and fuel. See Kumbhakar (1996) for a complete explanation of the data. Using a Cobb–Douglas specification, Horrace and Schmidt (1996) estimate the marginal products of capital, labor, and fuel to be: 0.5882, -0.0966, and 0.5807, respectively (only capital and fuel are significant at the 95% level). They also estimate $\hat{\sigma}_*$ = 0:0126. Ranked estimates of TE_j and $\hat{\mu}_{*j}$ are contained in Table 1. Notice TE_j is an increasing function of $\hat{\mu}_{*j}$ for fixed $\hat{\sigma}_*$ by Lemma 8.

Ignoring the sampling variability in $\hat{\mu}_{*j}$ and $\hat{\sigma}_*^2$ per Battese and Coelli (1988), Theorem 10 selects S_{min}, a subset of efficient firms (that have small u_j) with at least a

Table 1
Ranked Texas utilities

Firm	TE_j	$\hat{\mu}_{*j}$	$F^*_{\hat{\delta}^j}(\mathbf{0}, \hat{\boldsymbol{\mu}}_*, \hat{\sigma}_*^2\mathbf{I}_n)$	$F_{\hat{\delta}^j}(\mathbf{0}, \hat{\boldsymbol{\mu}}_*, \hat{\sigma}_*^2\mathbf{I}_n)$
5	0.9982	−0.0889	0.7117	0.0000
3	0.9960	−0.0325	0.2914	0.0000
10	0.9649	0.0358	0.0009	0.0000
1	0.9325	0.0700	0.0000	0.0000
8	0.9167	0.0870	0.0000	0.0000
9	0.8997	0.1058	0.0000	0.0011
2	0.8973	0.1084	0.0000	0.0019
6	0.8835	0.1239	0.0000	0.0285
7	0.8788	0.1293	0.0000	0.0607
4	0.8555	0.1561	0.0000	0.9074

confidence level of 1 - α. Assuming S_{min} is non-empty, Theorem 12 selects S_min, a subset of inefficient firms (that do not have small u_j) with at least a confidence level of 1- α. Let α = 0:10. Results for the Texas utilities are contained in Table 1.[6] The results for $F^*_{\#}$ are in the 4th column of the table. Notice that $F^*_{\#}$ is a decreasing function of $\tilde{\mu}_{*j}$. Based on the results the following conclusions can be drawn. First, S_{min} = $\emptyset$; so there is no inference on the single most efficient firm at the 90% confidence level. One can conclude that firm 5 is efficient with 71% probability and firm 3 is efficient with 29% probability, but these are not very strong inferential statements. Additionally, one might conclude that firms 3 or 5 (or both) are efficient with near certainty ð0:71 þ 0:29 ¼ 1Þ. Since S_{min} is empty, there is no guarantee that Theorem 12 will hold, but one can conclude that S_min = {10; 1; 8; 9; 2; 6; 7; 4} and that these firms are not most efficient with near certainty (0:71 + 0:29 = 1). $F_{\#}$ is calculated in column 5 of the table. Clearly, S_{max} = {4}, so firm 4 is least efficient with at least 90% confidence (in fact we are 90.74% confident). Since S_{max} is nonempty and S_max = {5; 3; 10; 1; 8; 9; 2; 6; 7}, one can conclude from Theorem 12 that these firms are not least efficient with at least 90% confidence (in fact we are 90.74% confident). A comparison to the inference of Horrace and Schmidt (1996, 2000) is in order. First, in Horrace and Schmidt (1996), their GLS specification calculates the same point estimates for TE_j as in Table 1. Their confidence intervals, based on critical points from univariate truncated normals, implicitly assume that technical efficiency is being measured relative to an unknown (out of sample) absolute standard. For instance, their 90% confidence interval for firm 5 is [0:9982 2 $\in$ 0:9721; 0:9994], so firm 5 is not operating on the absolutely efficient frontier with 90% probability. The inference presented here is for relative efficiency: firm 5 is efficient relative to the other firms with 71.17% probability. Of course, at the 90% level the inference determines that firm 5 may not be on the efficient frontier (Smin = $\emptyset$), so at least the results of the two different techniques confirm one another. One might conclude that with 90% probability firm 3 or 5 is the most efficient, 4 is least efficient, and the rest are somewhere in between. This a stronger statement than that of the Horrace and Schmidt (1996) intervals, which can only say that all the firms are absolutely inefficient.

Horrace and Schmidt (2000) calculate confidence intervals using a fixed-effect specification and ‘‘multiple comparisons with the best’’ techniques, based on differences of normal (non-truncated) variates. Their inference is explicitly for relative differences (similar to the results here) and imply a subset selection criterion for the relatively efficient firm. For instance, in that study firm 5 has a 90% confidence interval of [0.9448, 1.000], and the subset of firms that may be relatively efficient consists of firms 3 and 5. This is similar, but not identical, to the inference here, where at the 90% level no single firm is relatively efficient, but with near certainty firm 3 or 5 (or both) are efficient. Specification and distributional differences aside, the inferential differences are also driven by the fact that in Horrace and Schmidt (2000) efficiency is a time-invariant estimable parameter, while here efficiency is an unobserved error component that (potentially) varies with time. Estimating actual technical efficiency (not just its distribution) enables Horrace and Schmidt to identify a non-empty set, similar to Dunnett (1955).

6. Conclusions

This paper develops selection rules for performing inference on rankings of firmlevel technical efficiency in parametric stochastic frontier models. If we are willing to ignore the sampling variability in estimates of the mean and variance that underlie the truncated distributions that characterize technical inefficiency, then the suggested selection rules are a better gauge of inefficiency than the commonly used rankings of $\mathrm{E}(u_{jt}|\varepsilon_{jt})$, because the rules take into account the variance of the underlying distributions, while the conditional mean rankings do not. Other attempts at incorporating this variance into efficiency analysis have been made: Horrace and Schmidt (1996) use it to construct marginal confidence intervals for the conditional distribution of u, and Bera and Sharma (1999) use it as a proxy for production risk or uncertainty. However, neither one of these innovations is a substitute for the proposed selection rules, because neither exploits the multivariate distribution of the differences to draw inferences about who is technically efficient and who is not. If we are unwilling to ignore the sampling variability, then the conditional distribution of u is not necessarily truncated normal, and the power of the selection rule is suspect. However, so are the usual sample rankings of $\mathrm{E}(u_{jt}|\varepsilon_{jt})$, the confidence intervals of Horrace and Schmidt (1996), and virtually every application of parametric stochastic frontiers that provides firm-level technical efficiency rankings. Therefore understanding the nature of this sampling variability should be a high priority in the productivity research agenda. In

[6] Initially, the $F_{\#}$ and $F^*_{\#}$ were calculated for each firm using Simpson’s rule in the GAUSS programming language, however the integral would not converge due to the small value for $\hat{\sigma}_*$. This was not a problem with the probability integral itself, but with the tolerances for the intrinsic function in GAUSS for calculating the c.d.f. of a standard normal. Since this is only an exercise, the integration was calculated with rectangles (instead of trapezoids) with a width of 0.00001; however, the result should be viewed with caution due to the approximation.

the context of the selection rule, accommodation of the sampling variability would involve adjusting the power of the rule based on some quantification of the variability, but this problem is left for future research. Alternatively, the conditional distribution of u could be boot-strapped, then quantiles from the distribution of all differences could be simulated to perform inference, but this is no substitute for a well-developed distributional theory. Moreover, Bayesian approaches could be adopted that allow for ranking inference while viably controlling for sampling variability. Finally, it is interesting to speculate on theoretical and empirical extensions for the selection rules. Perhaps, they could be used for inference on truncated normal population parameters, based on random observations from the truncated populations. This seems reasonable, but the distributional theory may be cumbersome. Also, perhaps the rules could be adapted to allow ranking and selection of various econometric model specifications based on some positive acceptance criteria, such as R-squared or ''sum of squared errors'', insofar as these criteria possess positive distributions. This, however, remains to be seen.

Acknowledgements

Thanks to Ed Dudewicz, Jeff Racine, Duke Kao, Gary Koop, participants at the Current Developments in Productivity and Efficiency Measurement Conference, 2002, and three anonymous referees. Antoinette Metzger provided excellent editorial assistance. Research supported by the Office of the Vice Chancellor and Center for Policy Research, Syracuse University.

Appendix A

A.1. Mathematical appendix

Proof of Theorem 3 The characteristic function of W is

$$CF_{\mathbf{W}}(\mathbf{t},\mu,\mathbf{\Sigma}) = \mathrm{E}(e^{\imath\mathbf{t}'\mathbf{W}}) = \frac{\int_e^\infty \exp\{\imath\mathbf{t}'\mathbf{w}\}\,\exp\{-\frac{1}{2}(\mathbf{w}-\mu)'\mathbf{\Sigma}^{-1}(x-\mu)\}\,d\mathbf{w}}{\int_e^\infty \exp\{-\frac{1}{2}(\mathbf{w}-\mu)'\mathbf{\Sigma}^{-1}(\mathbf{w}-\mu)\}\,d\mathbf{w}}.$$

The proof proceeds with a series of variable transformations where the determinant of the Jacobian always equals 1. Let $\mathbf{G} = \mathbf{W} - \mu$, then $G \in [\mathbf{M}, \infty]$:

$$CF_{\mathbf{W}}(\mathbf{t},\mu,\mathbf{\Sigma}) = \frac{\int_M^\infty \exp\{\imath\mathbf{t}'\mathbf{g} - \frac{1}{2}(\mathbf{g}'\mathbf{\Sigma}^{-1}\mathbf{g})\}\,d\mathbf{g}}{\int_M^\infty \exp\{-\frac{1}{2}\mathbf{g}'\mathbf{\Sigma}^{-1}\mathbf{g}\}\,d\mathbf{g}}\exp\{\imath\mathbf{t}'\mu\}.$$

Now $\exp\{\imath\mathbf{t}'\mathbf{g} - \frac{1}{2}\mathbf{g}'\Sigma^{-1}\mathbf{g}\} = \exp\{-\frac{1}{2}\mathbf{t}'\Sigma\mathbf{t} - \frac{1}{2}(\mathbf{g}-\imath\Sigma\mathbf{t})'\Sigma^{-1}(\mathbf{g}-\imath\Sigma\mathbf{t})\}$ and

$$\begin{aligned} CF_{\mathbf{W}}(\mathbf{t},\mu,\mathbf{\Sigma}) &= \frac{\int_M^\infty \exp\{-\frac{1}{2}\mathbf{t}'\mathbf{\Sigma}\mathbf{t} - \frac{1}{2}(\mathbf{g}-\imath\mathbf{\Sigma}\mathbf{t})'\mathbf{\Sigma}^{-1}(\mathbf{g}-\imath\mathbf{\Sigma}\mathbf{t})\}\,d\mathbf{g}}{\int_M^\infty \exp\{-\frac{1}{2}\mathbf{g}'\mathbf{\Sigma}^{-1}\mathbf{g}\}\,d\mathbf{g}}\exp\{\imath\mathbf{t}'\mu\} \\ &= \frac{\int_M^\infty \exp\{-\frac{1}{2}(\mathbf{g}-\imath\mathbf{\Sigma}\mathbf{t})'\mathbf{\Sigma}^{-1}(\mathbf{g}-\imath\mathbf{\Sigma}\mathbf{t})\}\,d\mathbf{g}}{\int_M^\infty \exp\{-\frac{1}{2}\mathbf{g}'\mathbf{\Sigma}^{-1}\mathbf{g}\}\,d\mathbf{g}}\exp\{\imath\mathbf{t}'\mu - \tfrac{1}{2}\mathbf{t}'\mathbf{\Sigma}\mathbf{t}\}. \end{aligned}$$

Let $\mathbf{U} = \mathbf{G} - \imath\Sigma\mathbf{t}$, then $U \in [\mathbf{P}, \infty]$:

$$CF_{\mathbf{W}}(\mathbf{t},\mu,\mathbf{\Sigma}) = \frac{\int_P^\infty \exp\{-\frac{1}{2}\mathbf{u}'\mathbf{\Sigma}^{-1}\mathbf{u}\}\,d\mathbf{u}}{\int_M^\infty \exp\{-\frac{1}{2}\mathbf{u}'\mathbf{\Sigma}^{-1}\mathbf{u}\}\,d\mathbf{u}}\exp\{\imath\mathbf{t}'\mu - \tfrac{1}{2}\mathbf{t}'\mathbf{\Sigma}\mathbf{t}\}. \qquad \square$$

Proof of Lemma 8. Taking the partial derivative of E(Wj) with respect to μ in Eq. (1)

$$\frac{\partial \mathrm{E}(W_j)}{\partial \mu} = 1 - \frac{\mu_j}{\sigma_j}\left[\frac{\phi(\mu_j/\sigma_j)}{1-\Phi(\mu_j/\sigma_j)}\right] - \left[\frac{\phi(\mu_j/\sigma_j)}{1-\Phi(\mu_j/\sigma_j)}\right]^2$$
$$= \frac{V(W_j)}{\sigma_j^2} > 0.$$

Now, take the partial derivative w.r.t. σ_j:

$$\frac{\partial \mathrm{E}(W_j)}{\partial \sigma_j} = \frac{1}{\sigma_j}\left[\mathrm{E}(W_j) - \mu_j \frac{V(W_j)}{\sigma_j^2}\right],$$

Which is positive when $\mu_j \leqslant 0$. When $\mu_j > 0$:

$$\frac{\partial \mathrm{E}(W_j)}{\partial \sigma_j} = \frac{\mu_j}{\sigma_j}\left[\frac{\mathrm{E}(W_j)}{\mu_j} - \frac{V(W_j)}{\sigma_j^2}\right].$$

However, $\mathrm{E}(W_j)/\mu_j > 1$ and $0 < V(W_j)/\sigma_j^2 < 1$ so the difference is positive. Therefore $\partial \mathrm{E}(W_j)/\partial \sigma_j > 0$

References

Aigner, D.J., Lovell, C.A.K., Schmidt, P., 1977. Formulation and estimation of stochastic frontier production functions. Journal of Econometrics 6, 21–37.

Amemiya, T., 1974. Multivariate regression and simultaneous equations models when the dependent variables are truncated normal. Econometrica 42, 99–1012.

Battese, G.E., Coelli, T.J., 1988. Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. Journal of Econometrics 38, 387–399.

Battese, G.E., Coelli, T.J., 1992. Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India. Journal of Productivity Analysis 3, 153–170.

Beattie, A.W., 1962. Truncation in two variables to maximize a function of the means of a normal multivariate distribution. Australian Journal of Statistics 4, 1–3.

Bera, A.K., Sharma, S.C., 1999. Estimating production uncertainty in stochastic frontier models. Journal of Productivity Analysis 12, 187–210.

Cuesta, R.A., 2000. A production model with firm-specific temporal variation in technical efficiency: with application to Spanish dairy farms. Journal of Productivity Analysis 13, 139–158.

Dunnett, C.W., 1955. A multiple comparison procedure for comparing several treatments with a control. Journal of the American Statistical Society, 1096–1121.

Fernandez, C., Koop, G., Steel, M.F.J., 2002. Multiple output production with undesirable outputs: an application to nitrogen surplus in agriculture. Journal of the American Statistical Association 97, 432–442.

Geweke, J., 1991. Efficient simulation from the multivariate normal and Student t distributions subject to linear constraints. In: Kermidas, E.M., Kaufman, S.M. (Eds.), Computing Science and Statistics: Proceedings of the 23rd Symposium in the Interface, pp. 571–578. Interface Foundation of North America, Fairfax, VA.

Greene, W.H., 2005, Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. Journal of Econometrics (this issue).

Gupta, S.S., 1965. On some multiple decision (selection and ranking) rules. Technometrics 7, 225–245.

Heckman, J.J., 1976. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. Annals of Economic and Social Measurement 5/4, 475–492.

Hong, H., Shum, M., 2002. Econometric models of asymmetric ascending auctions. Journal of Econometrics 112, 327–358.

Horrace, W.C., 2003. Some Results on the Multivariate Truncated Normal Distribution. Center for Policy Research, Syracuse University, unpublished manuscript.

Horrace, W.C., Hernandez, C., 2001. The characteristic function from a family of truncated normal distributions. Econometric Theory, Problems & Solutions 17, 1026.

Horrace, W.C., Schmidt, P., 1996. Confidence statements for efficiency estimates from stochastic frontier models. Journal of Productivity Analysis 7, 257–282.

Horrace, W.C., Schmidt, P., 2000. Multiple comparisons with the best, with economic applications. Journal of Applied Econometrics 15, 1–26.

Jondrow, J., Lovell, C.A.K., Materov, I.S., Schmidt, P., 1982. On the estimation of technical efficiency in the stochastic production function model. Journal of Econometrics 19, 233–238.

Kim, Y., Schmidt, P., 2000. A review and empirical comparison of Bayesian and classical approaches to inference on efficiency levels in stochastic frontier models with panel data. Journal of Productivity Analysis 14, 91–118.

Koop, G., Osiewalski, J., Steel, M.F.J., 1997. Bayesian efficiency analysis through individual effects: hospital cost frontiers. Journal of Econometrics, 77–105.

Kumbhakar, S.C., 1990. Production frontiers, panel data, and time-varying technical inefficiency. Journal of Econometrics 46, 201–211.

Kumbhakar, S.C., 1996. Estimation of cost efficiency with heteroscedasticity: an application to electric utilities in Texas, 1966–1985. Journal of The Royal Statistical Society, Series D (The Statistician) 45, 319–335.

Rizvi, M.H., 1971. Some selection problems involving folded normal distributions. Technometrics 13, 355–369.

Tallis, G.M., 1961. The moment generating function of the truncated multi-normal distribution. Journal of the Royal Statistical Society, Series B 23, 223–229.

Tallis, G.M., 1963. Elliptical and radial truncation in normal populations. Annals of Mathematical Statistics 34, 940–944.

Tallis, G.M., 1965. Plane truncation in normal populations. Journal of the Royal Statistical Society, Series B 27, 301–307.

Tobin, J., 1958. Estimation of relationships for limited dependent variables. Econometrica 26, 24–36.

Tong, Y.L., 1990. The Multivariate Normal Distribution. Springer, New York.

Tsionas, E.G., 2002. Stochastic frontier models with random coefficients. Journal of Applied Econometrics 17, 127–147.

Weiler, H., 1959. Means and standard deviations of a truncated bivariate normal distribution. Australian Journal of Statistics 1, 73–81.