On Ranking and Selection from Independent Truncated Normal Distributions

William C. Horrace

Syracuse University and NBER

June, 2003

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 pre-specified probabilities. An empirical example is provided.

Key Words: Ranking and Selection, Truncated Normal, Stochastic Frontier.

JEL Code: C12, C16, C44, D24

On Ranking and Selection from Independent Truncated Normal Distributions

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 normals are used in the censored and truncated regression models; see Tobin (1958), Amemiya (1974) or Heckman (1976). Recently, Hong and Shum (2001) 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 (2002). 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 $\mathbf{W}' = [W_1, ..., W_n]$ is a multivariate truncated normal random variable, the goal is to attach probabilities to statements on the the relative magnitudes of the W_j when they are assumed independent. In particular, this paper presents probabilistic ranking and selection rules to determine subsets of the n elements of $\mathbf{W}$ that are relatively small or large at pre-specified probabilities. The proposed rules are based on a non-standard multivariate dis-

tribution 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 estimates 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 pre-specified probability, and it is argued that the rules are more theoretically justified than current methods for assessing distributional differences in these models.

The next section provides basic characterizations for the multivariate truncated normal distribution: density, characteristic function and moments. Section 3 provides results on linear transformations of independent truncated normal distributions. Section 4 provides the selection rules. Section 5 illustrates the main results with an empirical example: the efficiency of a panel of electric utilities in Texas. Section 6 summarizes and concludes.

2 Characterizations of the Distribution

This section presents characterizations of the multivariate truncated normal distribution.

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

$$f_{\mathbf{W}^*}(\mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-n/2} (\det \boldsymbol{\Sigma})^{-\frac{1}{2}} \exp\{-\frac{1}{2}(\mathbf{w} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{w} - \boldsymbol{\mu})\}; \mathbf{w} \in \Re^n.$$

Adopt the standard notation: $\mathbf{W}^* \sim N(\mathbf{0}, \boldsymbol{\Sigma})$. Let $\mathbf{W}' = [W_1, ..., W_n]$ be the truncation of $\mathbf{W}^*$ below $\mathbf{c} = [c_1, ..., c_n]$, $\mathbf{c} \in \Re^n$.

On Ranking and Selection from Independent Truncated Normal Distributions

Definition 2 W *has an n-variate zero-truncated normal distribution given by:*

$$
f_{\mathbf{W}}(\mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \;=\; \frac{(2\pi)^{-n/2}(\det \boldsymbol{\Sigma})^{-\frac{1}{2}}\exp\{-\frac{1}{2}(\mathbf{w}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{w}-\boldsymbol{\mu})\}}{(2\pi)^{-n/2}(\det \boldsymbol{\Sigma})^{-\frac{1}{2}}\int_{n}^{\infty}\exp\{-\frac{1}{2}(\mathbf{w}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{w}-\boldsymbol{\mu})\}d\mathbf{w}}; \mathbf{w} \in \Re_{\geq \mathbf{c}}^{n};
$$

$$
=\; \frac{\exp\{-\frac{1}{2}(\mathbf{w}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{w}-\boldsymbol{\mu})\}}{\int_{n}^{\infty}\exp\{-\frac{1}{2}(\mathbf{w}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{w}-\boldsymbol{\mu})\}d\mathbf{w}}; \mathbf{w} \in \Re_{\geq \mathbf{c}}^{n},
$$

where $\int_{n}^{\mathbf{b}}$ is a n-dimensional Riemann integral from $\mathbf{a}$ $(n \times 1)$ to $\mathbf{b}$ $(n \times 1)$, and $\Re_{\geq \mathbf{c}}^{n} = \{\mathbf{w} \in \Re^{n} : \mathbf{w} \geq \mathbf{c}\}$. 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). However, interest centers on parametric stochastic frontier models which produce distributions for firm-level technical efficiency that are described by the truncation in Definition 2. Define $(n \times 1)$ vectors: $\mathbf{M} = \mathbf{c} - \boldsymbol{\mu}$, $\mathbf{t} \in \Re^{n}$ and $\mathbf{P}$ $(n \times 1) = \mathbf{c} - \boldsymbol{\mu} - \iota\boldsymbol{\Sigma}\mathbf{t}$ with typical elements M_{j}, t_{j}, P_{j}; $j \in N$, $N = \{1, ..., n\}$, respectively and $\iota = \sqrt{-1}$. Then, the characteristic function of $\mathbf{W}$ is:

Theorem 3 *The characteristic function of* $\mathbf{W}$ *is given by:*

$$
CF_{\mathbf{W}}(\mathbf{t}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \;=\; \frac{\int_{n}^{\infty}\exp\{-\frac{1}{2}(\mathbf{w}-\iota\boldsymbol{\Sigma}\mathbf{t})'\boldsymbol{\Sigma}^{-1}(\mathbf{w}-\iota\boldsymbol{\Sigma}\mathbf{t})\}d\mathbf{w}}{\int_{n}^{\infty}\exp\{-\frac{1}{2}\mathbf{w}'\boldsymbol{\Sigma}^{-1}\mathbf{w}\}d\mathbf{w}}\exp\{\iota\mathbf{t}'\boldsymbol{\mu} - \frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}\}
$$

$$
=\; \frac{\int_{n}^{\infty}\exp\{-\frac{1}{2}\mathbf{u}'\boldsymbol{\Sigma}^{-1}\mathbf{u}\}d\mathbf{u}}{\int_{n}^{\infty}\exp\{-\frac{1}{2}\mathbf{u}'\boldsymbol{\Sigma}^{-1}\mathbf{u}\}d\mathbf{u}}\exp\{\iota\mathbf{t}'\boldsymbol{\mu} - \frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}\}.
$$

The proof is in the Mathematical Appendix. Tallis (1961) derives a similar formula for the moment generating function of $\mathbf{W}$. The derivation hinges on a series of variable transformations that shift $\mathbf{W}$ without scaling it. The univariate case $(n = 1)$ was first suggested in a problem posed by Horrace and Hernandez (2001). The characteristic function generates the moments of $\mathbf{W}$. Tallis (1961) derives these moments using differentiation of the

moment generating function when $\mathbf{W}^* \sim N(\mathbf{0}, \mathbf{R})$ and $\mathbf{R}$ is the correlation matrix associated with $\boldsymbol{\Sigma}$. Amemiya (1974) adapts the Tallis results for the case where $\mathbf{W}^* \sim N(\mathbf{0}, \boldsymbol{\Sigma})$. Weiler (1959) derives them for the case $n = 2$ using integration of the density $f_{\mathbf{W}}$. While most of the results of this paper can be derived for arbitrary truncation below $\mathbf{c}$, this paper is concerned with truncation below $\mathbf{0}$ for each element of $\mathbf{W}^*$. Moreover, ranking and selection rules are greatly simplified by orthogonality. Therefore, we will always make the following two assumptions.

Assumption A.0: $\mathbf{c} = \mathbf{0}$.

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 $\mathbf{W}^*$ are independent and $\mathbf{W}$ has an n-variate truncated normal distribution given by:*

$$f_{\mathbf{W}}(w_j, \mu_j, \sigma_j^2) = \frac{(2\pi)^{-n/2}\sigma^{-1}\exp\{-\frac{1}{2}\sum_{j=1}^{n}\frac{(w_j - \mu_j)^2}{\sigma_j}\}}{\prod_{j=1}^{n}[1 - \Phi(-\mu_j/\sigma_j)]}; w_j \in \Re_+,$$

where $\sigma^2 = \prod_{j=1}^{n}\sigma_j^2$ and $\Re_+^n = \{\mathbf{w} \in \Re^n : \mathbf{w} \geq \mathbf{0}\}$.

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 $\mathbf{W}$ is:*

$$F_{\mathbf{W}}(w_j, \mu_j, \sigma_j^2) = \frac{\prod_{j=1}^{n}[\Phi(\{w_j - \mu_j\}/\sigma_j) - \Phi(-\mu_j/\sigma_j)]}{\prod_{j=1}^{n}[1 - \Phi(-\mu_j/\sigma_j)]}; w_j \in \Re_+.$$

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 given by:*

$$CF_{\mathbf{W}}(t_j, \mu_j, \sigma_j^2) = \frac{\prod_{j=1}^{n}[1 - \Phi(-\mu_j/\sigma_j - \iota\sigma_j t_j)]}{\prod_{j=1}^{n}[1 - \Phi(-\mu_j/\sigma_j)]}\exp\{\sum_{j=1}^{n}(\iota t_j \mu_j - \frac{1}{2}t_j^2\sigma_j^2)\}.$$

4

On Ranking and Selection from Independent Truncated Normal Distributions

It is clear from Definition 4 or Corollary 6 that:

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

The result can be generalized to the case where W_j and W_i 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 to moments of the distribution are:

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

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:*

$$
\begin{array}{lll}
i) & E(W_j) = & \xi_j \text{ is increasing in } \mu_j \text{ , and} \\
ii) & E(W_j) = & \xi_j \text{ is increasing in } \sigma_j.
\end{array}
$$

The proof is in the Mathematical Appendix. Result i of Lemma 8 was indirectly shown by Bera and Sharma (1999). The implication is that if $\mu_j > \mu_i$ for fixed $\sigma_j = \sigma$ or if $\sigma_j > \sigma_i$ 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 equation 1, would be non-standard.

3   Linear Transformations

The selection procedures in the sequel hinge on the distributions of differences of truncated normals, so understanding the effects of linear transformations on the distribution is useful. First, without assumption A.1, the family of multivariate 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. One implication of this result is that the marginal distributions from multivariate truncated normals are not truncated normals. Second, even under assumption A.1, the family of multivariate normal distributions (Definition 4) is not closed to linear transformations. Summation (or differencing) of the distribution causes the distribution to loose its truncated normal shape. 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.) Therefore, the sampling distribution for the sample average from a random sample of a truncated normal population will not be truncated normal. However, the marginal distributions from independent truncated normals will be 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

---

[1] Proof of all of the preceding facts is available from the author.

difference $W_j - W_k$, $j \neq k$ is:

$$
f_{W_j - W_k}(w) = \begin{cases} \int_0^\infty f_{W_j}(x, \mu_j, \sigma_j^2)\, f_{W_k}(x - w, \mu_k, \sigma_k^2)dx & \text{for } w < 0 \\ \int_0^\infty f_{W_j}(w + x, \mu_j, \sigma_j^2)\, f_{W_k}(x, \mu_k, \sigma_k^2)dx & \text{for } w \geq 0 \end{cases} \tag{3}
$$

$$
= \begin{cases} \dfrac{1 - \Phi\left((\sigma_j^2 + \sigma_k^2)^{-1/2}\left[-\frac{\sigma_j}{\sigma_k}(w + \mu_k) - \mu_j \frac{\sigma_k}{\sigma_j}\right]\right)}{\{2\pi\sigma_j^2\sigma_k^2\}^{1/2}[1 - \Phi(-\mu_j/\sigma_j)][1 - \Phi(-\mu_k/\sigma_k)]} \exp\left\{-\dfrac{[w - (\mu_j - \mu_k)]^2}{2(\sigma_j^2 + \sigma_k^2)}\right\} & \text{for } w < 0 \\ \dfrac{1 - \Phi\left((\sigma_j^2 + \sigma_k^2)^{-1/2}\left[\frac{\sigma_k}{\sigma_j}(w - \mu_j) - \mu_k \frac{\sigma_j}{\sigma_k}\right]\right)}{\{2\pi\sigma_j^2\sigma_k^2\}^{1/2}[1 - \Phi(-\mu_j/\sigma_j)][1 - \Phi(-\mu_k/\sigma_k)]} \exp\left\{-\dfrac{[w - (\mu_j - \mu_k)]^2}{2(\sigma_j^2 + \sigma_k^2)}\right\} & \text{for } w \geq 0 \end{cases}
$$

where f_{W_j} is the marginal density function for W_j given in Definition 4 with $n = 1$. The partition of the integral on $w < 0$ and $w \geq 0$ is for computational convenience.[2] Notice that $W_j - W_k \in \Re$, so it 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 equation 3 to the $(n - 1)$-dimensional case where k is a control index. Define $\boldsymbol{\delta}^k \in \Re^{n-1}$ vector:

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

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

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

where F_{W_j} is the marginal distribution function for W_j given by Definition 5 with $n = 1$.

---

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

On Ranking and Selection from Independent Truncated Normal Distributions

The inverse distribution is:

$$
\begin{aligned}
F_{\boldsymbol{\delta}^k}^{-1}(\mathbf{d}, \mu_j, \sigma_j^2) &= \Pr\{\boldsymbol{\delta}^k \geq \mathbf{d}\} \quad (5)\\
&= \int_{\mathbf{W}:\, \delta^k \leq \mathbf{d}}^{n} f_{\mathbf{W}}(w, \mu_j, \sigma_j^2) dw\\
&= \int_0^\infty \int_{\substack{\max\{w_k+d_j,0\}\\ j\neq k}}^{\infty} \prod_{i=1}^{n} f_{W_i}(w_i, \mu_i, \sigma_i^2) dw\\
&= \int_0^\infty \prod_{\substack{j\neq k}}^{n} [1 - F_{W_j}(\max_{j\neq k}\{w_k + d_j, 0\}, \mu_j, \sigma_j^2)] f_{W_k}(w_k, \mu_k, \sigma_k^2) dw_k.
\end{aligned}
$$

The probabilities given in equations 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 $\alpha \in (0,1)$ and L_k^α as the solution in $d_j = d$ (for all j) to $F_{\boldsymbol{\delta}^k}^{-1}(d, \mu_j, \sigma_j^2) = 1 - \alpha$. Also, define $\gamma \in (0,1)$ and U_k^γ as the solution in $d_j = d$ (for all j) to $F_{\boldsymbol{\delta}^k}(d, \mu_j, \sigma_j^2) = \gamma$. Then, L_k^α and U_k^γ 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 $F_{\boldsymbol{\delta}^k}(d, \mu_j, \sigma_j^2) \neq 1 - F_{\boldsymbol{\delta}^k}^{-1}(d, \mu_j, \sigma_j^2)$.

4   Selection Rules

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

$$w_{[n]} > w_{[n-1]} > ... > w_{[1]}$$

be the ranks of the elements of a single potential realization of $\mathbf{W}$. Interest centers on selecting a subset of the indices $\{1, 2, ..., n\}$ that contains the index $[n]$ with a minimum pre-specified confidence level and another subset that contains $[1]$ with a minimum prespecified confidence level. Consider the following selection rules $R_{\max}$ and $R_{\min}$:

$$R_{\max} \quad : \quad \text{Select population } k \text{ if } U_k^{1-\gamma} \leq 0,$$

$$R_{\min} \quad : \quad \text{Select population } k \text{ if } L_k^{\alpha} \geq 0.$$

Furthermore, define corresponding subsets $S_{\max}$ and $S_{\min}$:

$$S_{\max} \quad = \quad \{k : U_k^{1-\gamma} \leq 0\}.$$

$$S_{\min} \quad = \quad \{k : L_k^{\alpha} \geq 0\}.$$

Notice that monotonicity of $F_{\boldsymbol{\delta}^k}^{-1}$ and $F_{\boldsymbol{\delta}^k}$ in d implies equivalent selection rules:

$$R_{\max} \quad : \quad \text{Select population } k \text{ if } 1 - \gamma \leq F_{\boldsymbol{\delta}^k}(0, \mu_j, \sigma_j^2) = \Pr\{[n] = k\} \tag{6}$$

$$R_{\min} \quad : \quad \text{Select population } k \text{ if } 1 - \alpha \leq F_{\boldsymbol{\delta}^k}^{-1}(0, \mu_j, \sigma_j^2) = \Pr\{[1] = k\}.$$

The continuity of $\boldsymbol{\delta}^k$ 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:

$$\Pr\{[n] \in S_{\max}|R_{\max}\} = \sum_{i \in S_{\max}} \Pr\{[n] = i\} = \sum_{i \in S_{\max}} F_{\boldsymbol{\delta}^i}(0, \mu_j, \sigma_j^2), \qquad (7)$$

$$\Pr\{[1] \in S_{\min}|R_{\min}\} = \sum_{i \in S_{\min}} \Pr\{[1] = i\} = \sum_{i \in S_{\min}} F_{\boldsymbol{\delta}^i}^{-1}(0, \mu_j, \sigma_j^2)$$

are valid. Of course, when $S_{\max}$ is empty, the probability given in equation 7 is just 0; similarly for $S_{\min}$. Let us always assume:

A.2: $\gamma < 0.5$ and $\alpha < 0.5$.

The following is a useful result.

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

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

Of course $S_{\max}$ and $S_{\min}$ can be empty, so there are only two states for the subsets: empty set or singleton. Given Lemma 9, equation 7 becomes:

$$\Pr\{[n] \in S_{\max}|R_{\max}\} = F_{\boldsymbol{\delta}^{(n)}}(0, \mu_j, \sigma_j^2) \qquad (8)$$

$$\Pr\{[1] \in S_{\min}|R_{\min}\} = F_{\boldsymbol{\delta}^{(1)}}^{-1}(0, \mu_j, \sigma_j^2)$$

where (n) and (1) correspond to the indices contained in $S_{\max}$ and $S_{\min}$, respectively, when the subsets are both singletons. That is, $S_{\max} = \{(n)\}$ and $S_{\min} = \{(1)\}$. This leads to the following result.

Theorem 10 *Let* $F_{\boldsymbol{\delta}^{(n)}}(0, \mu_j, \sigma_j^2) = 1 - \gamma_{(n)}$ *and* $F_{\boldsymbol{\delta}^{(1)}}^{-1}(0, \mu_j, \sigma_j^2) = 1 - \alpha_{(1)}$. *If* $S_{\max}$ *in non-empty, then the probability of a correct selection conditional on the selection rule* $R_{\max}$ *is:*

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

Similarly, if $S_{\min}$ *in non-empty, then the probability of a correct selection conditional on the selection rule* $R_{\min}$ *is:*

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

Otherwise, there is no inference.

Proof. The proof follows directly from equations 6 and 8 . ∎

For a prespecified confidence level $(1 - \gamma$ 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 maximum* W_j *and minimum* W_j, *respectively*. One could consider finding subsets containing indices with *low probability of corresponding to the maximum* W_j *and minimum* W_j. Therefore, alternative (but not equivalent) rules are:

$$R_{-\max} \quad : \quad \text{Select population } k \text{ if } U_k^{\gamma} \geq 0,$$

$$R_{-\min} \quad : \quad \text{Select population } k \text{ if } L_k^{1-\alpha} \leq 0,$$

with corresponding subsets $S_{-\max}$ and $S_{-\min}$. (Note the notational subtlety: "max" corresponds to "maximum with high probability", while "$-$max" corresponds to "maximum

---

[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.

with low probability" or "not the maximum".) Again, monotonicity implies:

$$R_{-\max} \quad : \quad \text{Select population } k \text{ if } \gamma \geq F_{\boldsymbol{\delta}^k}(0, \mu_j, \sigma_j^2)$$

$$R_{-\min} \quad : \quad \text{Select population } k \text{ if } \alpha \geq F_{\boldsymbol{\delta}^k}^{-1}(0, \mu_j, \sigma_j^2).$$

A useful result that relates selection rules and subsets is:

Lemma 11 *The sets $S_{\max}$ and $S_{-\max}$ are non intersecting. Also, the sets $S_{\min}$ and $S_{-\min}$ are non-intersecting.*

Proof. Under A.2, monotonicity of $F_{\boldsymbol{\delta}^k}$ implies $1 - \gamma > \gamma$ so that $U_k^{1-\gamma} \geq U_k^{\gamma}$. Therefore, if $R_{\max}$ selects k ($0 \geq U_k^{1-\gamma}$), then $R_{-\max}$ will not select i, because $0 \geq U_k^{1-\gamma} \geq U_k^{\gamma}$ violates the selection rule: $U_k^{\gamma} \geq 0$ (and vice versa). The proof is completed similarly for $S_{\min}$ 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:

$$\Pr\{[n] \quad \notin \quad S_{-\max}|R_{-\max}\} = \sum_{i \notin S_{\max}} \Pr\{[n] = i\} = \sum_{i \notin S_{\max}} F_{\boldsymbol{\delta}^i}(0, \mu_j, \sigma_j^2), \qquad (9)$$

$$\Pr\{[1] \quad \notin \quad S_{-\min}|R_{-\min}\} = \sum_{i \notin S_{\min}} \Pr\{[1] = i\} = \sum_{i \notin S_{\min}} F_{\boldsymbol{\delta}^i}^{-1}(0, \mu_j, \sigma_j^2).$$

These probabilities are not necessarily bound by the prespecified confidence level, unless $S_{\max}$ and $S_{-\max}$ are not empty.

Theorem 12 *If $S_{\max}$ is not empty, the probability of a correct selection conditional on the selection rule $R_{-\max}$ is $\Pr\{[n] \notin S_{-\max}|R_{-\max}\} \geq 1 - \gamma_{(n)} \geq 1 - \gamma$. If $S_{\min}$ is not empty, the probability of a correct selection conditional on the selection rule $R_{-\min}$ is $\Pr\{[1] \notin S_{-\min}|R_{-\min}\} \geq 1 - \alpha_{(1)} \geq 1 - \alpha$.*

Proof. When $S_{\max}$ is not empty,

$$\{[n] \in S_{\max}|R_{\max}\} \implies \{[n] \notin S_{-\max}|R_{-\max}\},$$

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

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

by Theorem 10. The proof is completed similarly for $S_{-\min}$. ∎

Of course, $\{[n] \notin S_{-\max}|R_{-\max}\}$ does not necessarily imply $\{[n] \in S_{\max}|R_{\max}\}$ so the $\Pr\{[n] \notin S_{-\max}|R_{-\max}\}$ may not be exactly $1 - \gamma_{(n)}$. In fact, the exact value is governed by equation 9. When $S_{\max}$ is empty, the probability of a correct selection is not bound from below and will be governed by equation 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, $n = 3$, $\mu_1 = 1$, $\mu_2 = 2$, $\mu_3 = 3$, $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 1$. Consider $F_{\boldsymbol{\delta}^k}(\mathbf{0}, \mu_j, \sigma_j^2)$.*

$$\begin{aligned}
F_{\boldsymbol{\delta}^1}(\mathbf{0}, \mu_j, \sigma_j^2) &= \Pr\{[n] = 1\} = 0.055 \\
F_{\boldsymbol{\delta}^2}(\mathbf{0}, \mu_j, \sigma_j^2) &= \Pr\{[n] = 2\} = 0.225 \\
F_{\boldsymbol{\delta}^3}(\mathbf{0}, \mu_j, \sigma_j^2) &= \Pr\{[n] = 3\} = 0.719 \\
F_{\boldsymbol{\delta}^1}^{-1}(\mathbf{0}, \mu_j, \sigma_j^2) &= \Pr\{[1] = 1\} = 0.696 \\
F_{\boldsymbol{\delta}^1}^{-1}(\mathbf{0}, \mu_j, \sigma_j^2) &= \Pr\{[1] = 2\} = 0.248 \\
F_{\boldsymbol{\delta}^1}^{-1}(\mathbf{0}, \mu_j, \sigma_j^2) &= \Pr\{[1] = 3\} = 0.056.
\end{aligned}$$

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.

Example 14 *Now suppose in the previous example that there is less variability and $\sigma_1^2 =$*

$\sigma_2^2 = \sigma_3^2 = 0.50.$ *Then*

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

$S_{\max}$ *is still empty, so none of the variables have high probability of being the maximum. However, $S_{-\max} = \{1\}$, since $F_{\delta^1} \leq \gamma$. Therefore, one can conclude that index 1 corresponds to the maximum with low probability. By equation 9 the probability of a correct selection conditional on $R_{-\max}$ is equal to $0.831 + 0.155 = 0.968$ which is happens to be greater than $1 - \gamma_{(n)} = 0.831$ and $1 - \gamma = .95$. However, Theorem 12 is not governing this high confidence level, because $S_{\max}$ is empty. Instead, the high confidence level is strictly an artifact of these particular distributional assumptions. Also, $S_{\min} = \emptyset$, 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 $\gamma = 0.05$, $S_{\max} = \emptyset$. However, $S_{-\max} = \{1, 2, 3\}$. By equation 9 the probability of a correct selection conditional on $R_{-\max}$ is equal to 0.663 which is less that $1 - \gamma$. This example demonstrates that the probability bounding of Theorem 12 requires $S_{\max}$ not be empty. Also, $S_{\min} = \emptyset$, 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 $\gamma = 0.05$, $S_{\max} = \{0\}$. 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 $S_{\max}$. Now $S_{-\max} = \{1, 2\}$. Therefore, one can conclude that indices 1 and 2 correspond to the maximum with low probability. By equation 9, the probability of a correct selection conditional on $R_{-\max}$ is $0.5 + 0.5 = 1.0$. Also, $S_{\min} = \emptyset$, 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 $\gamma = 0.05$, $S_{\max} = \{4\}$, so index 4 corresponds to the maximum with high probability. The probability of a correct selection conditional on $R_{\max}$ 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, $S_{\min} = \{1\}$, and $S_{-\min} = \{2, 3, 4\}$. ∎

It should be noted that all the probabilities in equations 7, 8, and 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 equations 4 and 5 is calculable, then theoretical implementations of these procedures may

be superior to simulation approaches.[4]

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 ineffi-

ciency that are independent normal random variables truncated below zero. Consider the

specification:

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

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 $v_{jt} \in \Re$ are random variables representing stochastic shocks

to the production process. Let the distribution of v_{jt} be that of an iid zero-mean normal

random variable with variance σ_v^2. Furthermore, let

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

where the $u_j \in \Re_+$ are positive random variables representing technical inefficiencies, and

$\gamma(t) > 0$ is some positive, continuous parameterization of t. Let the distribution of u_j be the

absolute value of an iid zero-mean 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

---

[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 at al. (2002), Tsionas (2002), Kim and Schmidt (2000), and Koop et al. (1997).

frontier" model has been extensively studied in various forms originating with Aigner, Lovell and Schmidt (1977) who restrict the model to $\gamma(t) = 1$, $T = 1$ 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 $g(x_{jt}, \beta) = x'_{jt}\beta$. 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 (2002) 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 $\epsilon_{jt} = v_{jt} - u_{jt}$ is a normal variable truncated below zero (this is also the case when u_{jt} is *iid* 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 a $N(\mu_{*j}, \sigma_*^2)$ variable where:

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

and

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

where $\bar{\epsilon} = T^{-1}\sum_{t=1}^{T}\epsilon_{jt}$. Additionally, in all the formulations neither realizations nor estimates of realizations of u_{jt} are available; only consistent estimates of the conditional mean and variance of u_{jt} 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 u_j is independent truncated normal, and an estimate of firm-level technical efficiency is the mean of the conditional truncated distribution: $E(u_j|\epsilon_{jt})$. Indeed, Battese and Coelli (1988, p391) state, "[w]e obtain the conditional distribution of the firm effect $[u_j]$, given the value of the random variables, $[\epsilon_{jt} = v_{jt} - u_j]$, $t = 1, 2, ..., T$. This assumes that the values of the parameter β are known." More recent formulations, based on time-varying technical inefficiency, suggest $E(u_{jt}|\epsilon_{jt})$ 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, $E(u_{jt}|\epsilon_{jt})$, over j in each period t as a proxy for the ranking of the unobserved random variable, u_{jt}. For example, see Horrace and Schmidt (1996). Unfortunately, $E(u_{jt}|\epsilon_{jt})$ is a misleading point estimate for u_{jt}. While smaller $E(u_{jt}|\epsilon_{jt})$ may suggest smaller u_{jt}, it may not be the case that u_{jt} is small in any particular sample, even if $E(u_{jt}|\epsilon_{jt})$ is small. Therefore, using $E(u_{jt}|\epsilon_{jt})$ as a point estimate of u_{jt} may have its limitations. Indeed, a firm j with $E(u_{jt}|\epsilon_{jt}) \simeq 0$ may be operating with $u_{jt} >> 0$ in any sample, y_{jt}, x_{jt}.

That being said, Theorems 10 or 12 (instead of rankings of the conditional means) are a better way to draw inferences on technical inefficiency. That is, use Theorem 10 to

---

[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.

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 $E(u_{jt}|\epsilon_{jt})$ 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 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 equation 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):

$$y_{jt} = x'_{jt}\beta + v_{jt} - u_j, \ j = 1, ..., n$$

Under the assumptions that $v_{jt} \sim iidN(0, \sigma_v^2)$; $u_j \sim |iidN(0, \sigma_u^2)|$; and x_{jt}, v_{tj}, u_j independent, Generalized Least Squares (GLS) yields consistent estimates σ_v^2 and σ_u^2 which imply the conditional distribution of u_j. Specifically, $\widehat{\mu}_{*j}$ and $\widehat{\sigma}_*^2$ are consistent for μ_{*j} and σ_*^2 in equations 11 and 12, respectively. Then the usual point estimates of technical efficiency

On Ranking and Selection from Independent Truncated Normal Distributions

based on Battese and Coelli (1988) are:

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

Horrace and Schmidt (1996) calculate the GLS technical efficiency of 10 Texas electric utility plants from a panel of data between 1966-1985, 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 $\widehat{\sigma}_* = 0.0126$. Ranked estimates of TE_j and $\widehat{\mu}_{*j}$ are contained in Table 1. Notice TE_j is an increasing function of $\widehat{\mu}_{*j}$ for fixed $\widehat{\sigma}_*$ by Lemma 8.

Ignoring the sampling variability in $\widehat{\mu}_{*j}$ and $\widehat{\sigma}_*^2$ per Battese and Coelli (1988), Theorem 10 selects $S_{\min}$, a subset of efficient firms (that have small u_j) with minimum $1-\alpha$ confidence confidence. Assuming $S_{\min}$ is non-empty, Theorem 12 selects $S_{-\min}$, a subset of inefficient firms (that have do not have small u_j) with minimum $1-\alpha$ confidence. Let $\alpha = 0.1$. Results for the Texas Utilities are contained in Table 1.[6] The results for $F_{\boldsymbol{\delta}^j}^{-1}$ are in the 4th column of the table. Notice that $F_{\boldsymbol{\delta}^j}^{-1}$ is a decreasing function of $\widehat{\mu}_{*j}$. It is unclear, whether this is a theoretical result or an artifact of the data, but it is certainly worth exploring in future research. Based on the results the following conclusions can be drawn. First, $S_{\min} = \emptyset$, so

---

[6] Initially, the $F_{\boldsymbol{\delta}^j}^+(\mathbf{0}, \widehat{\mu}_{*j}, \widehat{\sigma}_*)$ 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 $\widehat{\sigma}_*$. 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.

there is no inference on the most efficient firm. One *could* conclude that firm 5 is efficient with 71% probability and firm 3 is efficient with 29% probability, but these are not very useful inferential statements. 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_{δ^j} 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 not empty 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 and 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 $0.9982 \in [0.9721, 0.9994]$, 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 is not on the efficient frontier $(S_{\min} = \emptyset)$, so at least the results of the two different techniques confirm one another. One might conclude that with 90% probability either firm 3 or 5 is the most efficient, 4 is least efficient, and the rest are somewhere in between. This a stronger statement than one might get from 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 are relatively efficient consists of firms 3 and 5. This is in contrast to the inference here, where at the 90% level no firm is relatively efficient. Specification and distributional differences aside, the inferential differences are 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 firm-level 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 ranking of $E(u_{jt}|\epsilon_{jt})$, 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 $E(u_{jt}|\epsilon_{jt})$, 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 stochastic frontier research agenda. In 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 truncated distributions. This, however, also remains to be seen.

7 Bibliography

1. Aigner, D.J., C.A.K. Lovell And P. Schmidt (1977). Formulation and Estimation of Stochastic Frontier Production Functions. *Journal of Econometrics*, 6, 21-37.

2. Amemiya (1974) Multivariate Regression and Simultaneous Equations Models When the Dependent Variables Are Truncated Normal, *Econometrica*, 42, 99-1012.

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

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

5. 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.

6. Bera, A.K. and S.C. Sharma (1999) Estimating Production Uncertainty in Stochastic Frontier Models. *Journal of Productivity Analysis*, 12, 187-210.

7. 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, 139-58.

8. Dunnett, C.W. (1955) A Multiple Comparison Procedure for Comparing Several Treatment with a Control, *Journal of the American Statistical Society*, 1096-1121.

9. Fernandez, C., G. Koop and M.F.J. Steel (2002). Multiple Output Production with Undesirable Outputs: An Application to Nitrogen Surplus in Agriculture, *Journal of the American Statistical Association*, 97, 432-442.

10. Geweke, J. (1991) Efficient Simulation from the Multivariate Normal and Student-t Distributions Subject to Linear Constraints, in *Computing Science and Statistics: Proceedings of the 23rd Symposium in the Interface*, E.M. Kermidas and S.M. Kaufman (eds.). Interface Foundation of North America.

11. Greene, W.H. (2002) Alternative Estimators for Stochastic Frontier Models, Unpublished Manuscript, Department of Economics, Stern Business School, New York University.

12. Gupta, S.S. (1965) On Some Multiple Decision (Selection and Ranking) Rules, *Tech-*

nometrics, 7, 225-245.

13. 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

14. Horrace, W.C. and C. Hernandez (2001) The Characteristic Function from a Family of Truncated Normal Distributions, Econometric Theory, Problems & Solutions, 17, 1026.

15. Horrace, W.C. and P. Schmidt (1996) Confidence Statements for Efficiency Estimates from Stochastic Frontier Models. *Journal of Productivity Analysis*, 7, 257-82.

16. Horrace, W.C. and P. Schmidt (2000) Multiple Comparisons with the Best, with Economic Applications. *Journal of Applied Econometrics*, 15, 1-26.

17. Hong, H. and M. Shum (2001) Econometrics Models of Asymmetric Ascending Auctions, Unpublished Manuscript, Department of Economics, Princeton University.

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

19. Kim Y. and P. Schmidt (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.

20. Koop G., J. Osiewalski and M.F.J. Steel (1997). Bayesian Efficiency Analysis Through Individual Effects: Hospital Cost Frontiers, *Journal of Econometrics*, 77-105.

21. Kumbhakar (1990). Production frontiers, panel data, and time-varying technical inefficiency. *Journal of Econometrics*, 46, 201-11.

22. Kumbhakar (1994). Estimation of Cost Efficiency with Heteroscedasticity: an Application to Electric Utilities in Texas, 1966-1985. *Journal of The Royal Statistical Society, Ser. D*, 45, 319-335.

23. Rizvi, M.H. (1971) Some Selection Problems Involving Folded Normal Distributions, *Technometrics*, 13, 355-369.

24. 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.

25. Tallis, G.M. (1963) Elliptical and Radial Truncation in Normal Populations, *Annals of Mathematical Statistics*, 34, 940-944.

26. Tallis, G.M. (1965) Plane Truncation in Normal Populations, *Journal of the Royal Statistical Society, Series B*, 27, 301-307.

27. Tong, Y.L. (1990) *The Multivariate Normal Distribution*, New York: Springer-Verlag.

28. Tobin, J. (1958) Estimation of Relationships for Limited Dependent Variables, *Econometrica*,26, 24-36.

29. Tsionas, E.G. (2002) Stochastic Frontier Models with Random Coefficients, *Journal of Applied Econometrics*, 17, 127- 47.

30. Weiler, H. (1959) Means and Standard Deviations of a Truncated Bivariate Normal Distribution, *Australian Journal of Statistics*, 1, 73-81

Tables

Table 1. Ranked Texas Utilities.

Firm	TE_j	$\widehat{\mu}_{*j}$	$F_{\boldsymbol{\delta}^j}^{-1}(\mathbf{0}, \widehat{\mu}_{*j}, \widehat{\sigma}_*)$	$F_{\boldsymbol{\delta}^j}(\mathbf{0}, \widehat{\mu}_{*j}, \widehat{\sigma}_*)$
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

Appendix A

A   Mathematical Appendix

Proof of Theorem 3.   The CF of W is:

$$CF_{\mathbf{W}}(\mathbf{t}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = E(e^{\iota \mathbf{t}' \mathbf{W}}) = \frac{\int_{\overset{n}{\mathbf{c}}}^{\infty} \exp\{\iota \mathbf{t}' \mathbf{w}\} \exp\{-\frac{1}{2}(\mathbf{w} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(x - \boldsymbol{\mu})\} d\mathbf{w}}{\int_{\overset{n}{\mathbf{c}}}^{\infty} \exp\{-\frac{1}{2}(\mathbf{w} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(\mathbf{w} - \boldsymbol{\mu})\} d\mathbf{w}}.$$

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

$$CF_{\mathbf{W}}(\mathbf{t}, \mu, \boldsymbol{\Sigma}) = \frac{\int_{\overset{n}{M}}^{\infty} \exp\{\iota \mathbf{t}' \mathbf{g} - \frac{1}{2}(\mathbf{g}' \boldsymbol{\Sigma}^{-1} \mathbf{g})\} d\mathbf{g}}{\int_{\overset{n}{M}}^{\infty} \exp\{-\frac{1}{2} \mathbf{g}' \boldsymbol{\Sigma}^{-1} \mathbf{g}\} d\mathbf{g}} \exp\{\iota \mathbf{t}' \boldsymbol{\mu}\}.$$

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

$$\begin{aligned}
CF_{\mathbf{W}}(\mathbf{t}, \mu, \boldsymbol{\Sigma}) &= \frac{\int_{\overset{n}{M}}^{\infty} \exp\{-\frac{1}{2} \mathbf{t}' \Sigma \mathbf{t} - \frac{1}{2}(\mathbf{g} - \iota \Sigma \mathbf{t})' \boldsymbol{\Sigma}^{-1}(\mathbf{g} - \iota \Sigma \mathbf{t})\} d\mathbf{g}}{\int_{\overset{n}{M}}^{\infty} \exp\{-\frac{1}{2} \mathbf{g}' \boldsymbol{\Sigma}^{-1} \mathbf{g}\} d\mathbf{g}} \exp\{\iota \mathbf{t}' \boldsymbol{\mu}\} \\
&= \frac{\int_{\overset{n}{M}}^{\infty} \exp\{-\frac{1}{2}(\mathbf{g} - \iota \Sigma \mathbf{t})' \boldsymbol{\Sigma}^{-1}(\mathbf{g} - \iota \Sigma \mathbf{t})\} d\mathbf{g}}{\int_{\overset{n}{M}}^{\infty} \exp\{-\frac{1}{2} \mathbf{g}' \boldsymbol{\Sigma}^{-1} \mathbf{g}\} d\mathbf{g}} \exp\{\iota \mathbf{t}' \boldsymbol{\mu} - \frac{1}{2} \mathbf{t}' \Sigma \mathbf{t}\}.
\end{aligned}$$

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

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

∎

Proof of Lemma 8.   Taking the partial derivative of $E(W_j)$ with respect to μ in equation 1:

$$\begin{aligned}
\frac{\partial E(W_j)}{\partial \mu} &= 1 - \frac{\mu_j}{\sqrt{\sigma_{jj}}} \left[\frac{\phi\left(\mu_j / \sqrt{\sigma_{jj}}\right)}{1 - \Phi\left(\mu_j / \sqrt{\sigma_{jj}}\right)} \right] - \left[\frac{\phi\left(\mu_j / \sqrt{\sigma_{jj}}\right)}{1 - \Phi\left(\mu_j / \sqrt{\sigma_{jj}}\right)} \right]^2 \\
&= \frac{V(W_j)}{\sigma_{jj}} > 0.
\end{aligned}$$

Now, taking the partial derivative w.r.t. σ:

$$\frac{\partial E(W_j)}{\partial \sigma} = \frac{1}{\sqrt{\sigma_{jj}}} \left[E(W_j) - \mu_j \frac{V(W_j)}{\sigma_{jj}} \right]$$

which is positive when $\mu_j \leq 0$. When $\mu_j > 0$:

$$\frac{\partial E(W_j)}{\partial \sigma} = \frac{\mu_j}{\sqrt{\sigma_{jj}}} \left[\frac{E(W_j)}{\mu_j} - \frac{V(W_j)}{\sigma_{jj}} \right].$$

However, $\frac{E(W_j)}{\mu_j} > 1$ and $0 < \frac{V(W_j)}{\sigma_{jj}} < 1$, so the difference is positive. Therefore, $\frac{\partial E(W_j)}{\partial \sigma} > 0$.

∎