

An Information Theoretic Approach to Flexible Stochastic Frontier Models

Douglas J. Miller*

July 16, 2007

Abstract

Parametric stochastic frontier models have a long history in applied production economics, but the class of tractible parametric models is relatively small. Consequently, researchers have recently considered non-parametric alternatives such as kernel density estimators, functional approximations, and data envelopment analysis (DEA). The purpose of this paper is to present an information theoretic approach to constructing more flexible classes of parametric stochastic frontier models. Further, the proposed class of models nests all of the commonly used parametric methods as special cases, and the proposed modeling framework provides a comprehensive means to conduct model specification tests. The modeling framework is also extended to develop information theoretic measures of mean technical efficiency and to construct a profile likelihood estimator of the stochastic frontier model.

Keywords: Kullback–Leibler information criterion, output distance function, profile likelihood, stochastic frontier, technical efficiency

JEL Classifications: C13, C21, C51

*118 Professional Building, Department of Economics, University of Missouri, Columbia, MO 65211

1 Introduction

Stochastic frontier models of production systems were originally proposed by Aigner, Lovell, Schmidt (1977) and by Meeusen and van den Broeck (1977). For the single output case, the general form of the stochastic frontier model is

$$y = g(\mathbf{x}) \exp(v - u) \tag{1}$$

where y is the observed production output, $g(\mathbf{x})$ is the conditional mean function given inputs \mathbf{x} , v is assumed to be a mean-zero error term that represents measurement error in the output level noise or other types of idiosyncratic noise, and u is a firm-specific random effect that represents the firm's technical inefficiency. As such, the production relationship has a multiplicative error process composed of $v \in \mathbb{R}$ and $u \in \mathbb{R}_+$. Assuming $E[\exp(v) | \mathbf{x}] = 1$, the conditional mean of y (given \mathbf{x}) is $E[y | \mathbf{x}] \equiv h(\mathbf{x}) \leq g(\mathbf{x})$ and the mean technical efficiency (across firms) is measured by $E[\exp(-u) | \mathbf{x}] \leq 1$ where the inequality is strict for some \mathbf{x} under technically inefficient production. For example, the relationship between $h(\mathbf{x})$ and $g(\mathbf{x})$ under technically inefficient production is illustrated in Figure 1. Under technical efficiency, $E[\exp(-u) | \mathbf{x}] = 1$ and the conditional mean achieves the production frontier such that $g(\mathbf{x}) = h(\mathbf{x})$ for all \mathbf{x} .

Parametric stochastic frontier models have a long history in applied production economics, but the class of tractible parametric models is relatively small. For example, the conditional mean (regression) function is typically specified as a traditional production function (e.g., Cobb–Douglas or translog), the real-valued error component v is commonly represented as a $N(0, \sigma_v^2)$ random variable, and the model for the positive-valued error component u is commonly selected from the following set

- Exponential: $u \sim \text{Exp}(\sigma_u)$ for $\sigma_u > 0$
- Half-normal: $u \sim N_+(0, \sigma_u^2)$ for $\sigma_u^2 > 0$
- Truncated normal: $u \sim N_+(\mu, \sigma_u^2)$ for $\mu \in \mathbb{R}$ and $\sigma_u^2 > 0$
- Gamma: $u \sim \text{Gamma}(m, \sigma_u)$ for $m > -1$ and $\sigma_u > 0$

Aigner and Chu (1968) presented the first non-stochastic frontier model (i.e., $v = 0$), and they used the exponential model to capture negative deviations from the frontier. Aigner, Lovell, and Schmidt (1977) adopted the exponential form for their seminal work on the stochastic frontier model, and they also introduced the half-normal model along with Meeusen and van den Broeck (1977). The exponential and half-normal models have modes at zero, which may be unrealistic for many production applications. Stevenson (1980) proposed the truncated normal model to allow for a positive mode in the distribution of u . He also introduced the gamma model, which was extended by

Greene (1980a, 1990). Further, the truncated normal model reduces to the half-normal model if $\mu = 0$ and the gamma model reduces to the exponential model if $m = 0$, and these relationships are evident in the first two moments of the candidate models (see Table 1). Examples of the candidate models are presented in Figure 2.

Estimators of the model parameters based on the maximum likelihood, method of moments, or related estimation criteria have the usual statistical properties if the models are correctly specified, but these properties depend critically on the assumed model specification. For these reasons, researchers have recently explored more robust non-parametric alternatives such as kernel density estimators and functional approximation methods as well as mathematical programming procedures like data envelopment analysis (DEA) to avoid model specification problems. In general, the non-parametric methods are consistent and perhaps asymptotically efficient, but the convergence rates are typically slower than in the parametric case and very large samples are required to avoid practical problems associated with smoothing parameter selection, polynomial order selection, or local data sparsity. DEA has become a very widely used tool in the applied production economics literature because it is also a model-free approach based on a relatively simple linear programming algorithm. DEA also provides a relatively tractable means to analyze multi-output production data. However, DEA cannot be used to directly evaluate many behavioral hypotheses that can be readily tested in the parametric context, and it may exhibit weaker performance than parametric stochastic frontier models that have roughly correct specifications (Gong and Sickles, 1989 and 1992). Consequently, there may be some potential gains from developing a more robust set of parametric stochastic frontier models.

The purpose of this paper is to present an information theoretic approach to constructing more flexible classes of parametric stochastic frontier models for single-output production systems. The proposed approach retains all of the favorable estimation and hypothesis testing properties associated with parametric models while providing a more flexible stochastic frontier specification. Further, the proposed class of models nests all of the commonly used parametric methods as special cases, and the proposed modeling framework provides a convenient means to conduct model specification tests. The rest of the paper is organized as follows: the information theoretic approach to modeling density and regression functions is outlined in Section 2. These methods are then used to construct a stochastic frontier model that nests the exponential, half-normal, truncated normal, and gamma models as special cases in Section 3. Measures of firm-specific and mean technical efficiency are highlighted in Section 4, and an extension of the information theoretic approach to profile likelihood models of the stochastic frontier are presented in Section 5. Concluding remarks are provided in Section 6.

2 Entropy, Density, and Regression Functions

The information theoretic approach to deriving probability density and mass functions from moments and other properties of random variables was initially developed by Jaynes (1957a, b). The maximum entropy modeling procedure and its extension to regression functions are reviewed in this section.

2.1 Maximum entropy density functions

Jaynes' maximum entropy method is based on an information theoretic measure of uncertainty known as Shannon's entropy functional. For a random variable y with probability density or mass function $f(y)$ and cumulative distribution function $F(y)$, Shannon's entropy is

$$H(f) = - \int \ln(f(y)) dF(y) \quad (2)$$

The entropy functional measures the uncertainty expressed in $f(y)$ about possible outcomes of random variable y . For discrete random variables with K possible outcomes, the Shannon entropy of a degenerate distribution is $H(f) = 0$ (i.e., there is no uncertainty about outcomes of y), and $H(f)$ increases to $\ln(K)$ as $f(y)$ becomes more diffuse on the support of y (i.e., $f(y)$ approaches a discrete uniform distribution with probabilities K^{-1}). For continuous random variables, the lower bound may not be zero, but $H(f)$ does increase as $f(y)$ becomes more diffuse on the support of y . For example, $H(f) = 1/2 + \ln(\sigma\sqrt{2\pi})$ for $y \sim N(\mu, \sigma^2)$. The uncertainty about outcomes of y is not affected by location shifts (i.e., changes in μ), but the uncertainty about outcomes of y increases with the variance of the random variable, σ^2 .

The maximum entropy method is designed to derive probability distributions with given moments or other properties, and the procedure selects f to maximize Shannon's entropy subject to these conditions. Following Jaynes' motivation for the method, the resulting probability distributions satisfy the known conditions but are otherwise "most uncertain" or "least informative" with regard to other properties of y . For example, suppose we know that some continuous random variable y has real-valued support $y \in \mathbb{R}$, mean $E[y] = \mu$, and variance $\text{var}(y) = \sigma^2$ but we do not know the probability distribution of y . Under Jaynes' method of maximum entropy, the distribution may be recovered by maximizing

$$H(f) = - \int_{-\infty}^{\infty} f(y) \ln(f(y)) dy \quad (3)$$

subject to

$$\mu = \int_{-\infty}^{\infty} y f(y) dy \quad (4)$$

$$\mu^2 + \sigma^2 = \int_{-\infty}^{\infty} y^2 f(y) dy \quad (5)$$

$$1 = \int_{-\infty}^{\infty} f(y) dy \quad (6)$$

where the first two constraints are the moment conditions and the third constraint is the required additivity property of $f(y)$. By the calculus of variations, the Lagrange expression is

$$\begin{aligned} \mathcal{L}(f, \gamma) = & - \int_{-\infty}^{\infty} f(y) \ln(f(y)) dy + \gamma_1 \left[\mu - \int_{-\infty}^{\infty} y f(y) dy \right] \\ & + \gamma_2 \left[\mu^2 + \sigma^2 - \int_{-\infty}^{\infty} y^2 f(y) dy \right] + \gamma_0 \left[1 - \int_{-\infty}^{\infty} f(y) dy \right] \end{aligned} \quad (7)$$

where γ_1 and γ_2 are the Lagrange multipliers on the moment constraints (4) and (5) and γ_0 is the Lagrange multiplier for the additivity constraint (6). The associated Euler conditions are

$$-\ln(\hat{f}(y)) - 1 - \hat{\gamma}_1 y - \hat{\gamma}_2 y^2 - \hat{\gamma}_0 = 0 \quad (8)$$

and the intermediate solution to the problem takes the form

$$\hat{f}(y) \propto \exp(-y\hat{\gamma}_1 - y^2\hat{\gamma}_2) \quad (9)$$

where $\hat{\gamma}_1$ and $\hat{\gamma}_2$ are the optimal Lagrange multipliers on constraints (4) and (5). The intermediate solution may be further developed to show that $\hat{\gamma}_1 = -\mu/\sigma^2$ and $\hat{\gamma}_2 = (2\sigma^2)^{-1}$, and the solution to the problem is the $N(\mu, \sigma^2)$ density function. Thus, the most uniform (most uncertain or least informative) distribution on the real line with known mean and variance is the normal distribution. The uniform, beta, exponential, gamma, Laplace (double exponential), and other familiar parametric probability distributions may also be derived as maximum entropy distributions under appropriate conditions, and further details are provided by Golan, Judge, and Miller (1996); Maasoumi (1993); Ryu (1993); Theil and Fiebig (1984); and Zellner and Highfield (1998). Miller and Liu (2002) extend this approach to derive multivariate density functions from given joint moment conditions (e.g., correlation or covariance) and compare the properties of the joint entropy-based distributions to copula-based probability models.

2.2 Maximum entropy regression functions

Many of the regression function models commonly used in applied economic research may also be derived as maximum entropy outcomes, and this approach follows the im-

portant contributions of Maasoumi (1993) and Ryu (1993). In the context of a single-output production system, this approach is based on the assumption that the output (y) and input (\mathbf{x}) variables are jointly distributed random variables. Accordingly, the joint probability density function $f(y, \mathbf{x})$ represents the simultaneously determined character of the input and output variables across the range of possible outcomes (conditional on input and output prices, technology, weather, etc.). Further, the conditional mean (regression) function $E[y | \mathbf{x}] \equiv h(\mathbf{x})$ may be interpreted as the kernel of an associated density function. To see this point, note that the unconditional expected value of y may be expressed as

$$\mu_y = \int h(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} = \int \int y f(y | \mathbf{x}) f(\mathbf{x}) dy d\mathbf{x} \quad (10)$$

where $f(y | \mathbf{x})$ is the conditional density function of y given \mathbf{x} , $f(\mathbf{x})$ is the marginal density function of \mathbf{x} , and μ_y is the unconditional expected value of y . Equation (10) implies that the following function of \mathbf{x}

$$\varphi_h(\mathbf{x}) = \frac{h(\mathbf{x}) f(\mathbf{x})}{\mu_y} \geq 0 \quad (11)$$

has the same properties as a density function (i.e., weakly positive and integrates to one over the support of \mathbf{x}). Accordingly, a second probability density function based on the production frontier may be defined as

$$\varphi_g(\mathbf{x}) = \frac{g(\mathbf{x}) f(\mathbf{x})}{\mu_y^e} \geq 0 \quad (12)$$

where μ_y^e represents the unconditional mean of efficient output. In the absence of technical inefficiencies, the density functions coincide such that $h(\mathbf{x}) = g(\mathbf{x})$ and $\varphi_h(\mathbf{x}) = \varphi_g(\mathbf{x})$ for all \mathbf{x} (almost everywhere).

Accordingly, Maasoumi (table 3, 1993) and Ryu (table 1, 1993) show that many of the regression functions $g(\mathbf{x}, \boldsymbol{\beta})$ commonly used in practice can be derived as maximum entropy outcomes under appropriate conditions on the relationship between the output level (y) and the inputs (\mathbf{x}). In particular, the class of maximum entropy regression functions includes the Cobb–Douglas, translog, generalized Cobb–Douglas, generalized Leontief, and miniflex–Laurent production functions plus the Fourier flexible form. Further, the joint density modeling approach presented by Miller and Liu (2002) may be used to derive the translog function used in multi-output stochastic frontier models (e.g., see equation (2) in Sickles, Good, and Getachew (2002)) as a maximum entropy functional form. Consequently, the maximum entropy criterion may be used to derive the regression model components $g(\mathbf{x})$ for stochastic frontier analysis.

3 General Class of Stochastic Frontier Models

Based on the results in Section 2.1, the $N(0, \sigma_v^2)$ model for v may be motivated as a maximum entropy specification — it is the least informative distribution supported on the real line with mean zero and variance σ_v^2 . The candidate models for u may also be derived as maximum entropy distributions on the non-negative real line under the moment constraints stated in Table 2 (plus the additivity constraint). As with the moments stated in Table 1, the relationships among the candidate models may be linked through the Lagrange multipliers. For example, the truncated normal model reduces to the half-normal model if $\gamma_1 = 0$ and the gamma model reduces to the exponential model if $\gamma_2 = 0$.

3.1 Deriving the general frontier model

More importantly, the maximum entropy approach presented in Section 2 may be extended to construct a general probability model for u based on the full set of constraints on y stated in Table 2. Specifically, the general frontier model is derived by maximizing

$$H(f) = - \int_{-\infty}^{\infty} f(u) \ln(f(u)) du \quad (13)$$

subject to

$$\mu = \int_0^{\infty} u f(u) du \quad (14)$$

$$\mu^2 + \sigma_u^2 = \int_0^{\infty} u^2 f(u) du \quad (15)$$

$$\psi = \int_0^{\infty} \ln(u) f(u) du \quad (16)$$

$$1 = \int_0^{\infty} f(u) du \quad (17)$$

where $\psi \equiv E[\ln(u)]$. The resulting model is the least informative distribution for u on the non-negative real line subject to the constraints in (14)–(16). By the calculus of variations, the Lagrange expression is

$$\begin{aligned} \mathcal{L}(f, \gamma) = & - \int_0^{\infty} f(u) \ln(f(u)) du + \gamma_1 \left[\mu - \int_0^{\infty} u f(u) du \right] \\ & + \gamma_2 \left[\mu^2 + \sigma^2 - \int_0^{\infty} u^2 f(u) du \right] + \gamma_3 \left[\psi - \int_0^{\infty} \ln(u) f(u) du \right] \\ & + \gamma_0 \left[1 - \int_0^{\infty} f(u) du \right] \end{aligned} \quad (18)$$

and the associated Euler conditions are

$$-\ln(\hat{f}(u)) - 1 - \hat{\gamma}_1 u - \hat{\gamma}_2 u^2 - \hat{\gamma}_3 \ln(u) - \hat{\gamma}_0 = 0 \quad (19)$$

The intermediate solution to the maximum entropy problem takes the form

$$\begin{aligned}\widehat{f}(u) &\propto \exp(-u\widehat{\gamma}_1 - u^2\widehat{\gamma}_2 - \ln(u)\widehat{\gamma}_3) \\ &= u^{-\widehat{\gamma}_3} \exp[-u\widehat{\gamma}_1 - u^2\widehat{\gamma}_2]\end{aligned}\tag{20}$$

which is the kernel function for the general frontier model

$$f(u) = \frac{u^{-\gamma_3} \exp(-u\gamma_1 - u^2\gamma_2)}{J_u(\boldsymbol{\gamma})}\tag{21}$$

parameterized by the Lagrange multipliers γ_1 , γ_2 , and γ_3 where $J_u(\boldsymbol{\gamma})$ is the normalizing constant. Note that the model is not directly parameterized in terms of the moment constraints (μ , σ_u^2 , and ψ), but the Lagrange multipliers are indirectly linked to the values of these moments. Accordingly, the proposed probability model (21) may be viewed as a member of the exponential family of distributions parameterized in canonical form. Further, the integral required to compute $J_u(\boldsymbol{\gamma})$ is similar in form to an integral in the likelihood function for the normal-gamma stochastic frontier model, and a closed-form expression for $J_u(\boldsymbol{\gamma})$ is reported by Beckers and Hammond (1987).

For comparison purposes, one member of the general frontier class is shown with a special case (gamma PDF with the same mode) in Figure 3. More importantly, we can show that the general frontier model also encompasses several other probability models, and the particular special cases are presented in Table 3. Most of these alternative models have properties appropriate for $f(u)$, but they have not been used in the existing stochastic frontier literature. Consequently, the proposed approach to deriving general frontier models allows us to consider a much broader class of alternative parametric models. Also, please note that the class of alternative models nested within the general frontier model may be further expanded by including additional moment constraints (e.g., third non-central moment of y) in the maximum entropy problem.

3.2 Estimating the general frontier model

To demonstrate the implicit linkage between the moment conditions for the maximum entropy problem and the Lagrange multipliers, the general model from (21) may be substituted back into the Lagrange equation (18). The resulting expression reduces to the concentrated objective function

$$\mathcal{M}(\boldsymbol{\gamma}) = \gamma_1 E[u] + \gamma_2 \text{var}(u) + \gamma_3 E[\ln(u)] + \ln(J_u(\boldsymbol{\gamma}))\tag{22}$$

By the saddle point property of the maximum entropy problem, this function is strictly convex in $\boldsymbol{\gamma}$, and the optimal values of the Lagrange multipliers may be determined by minimizing $\mathcal{M}(\boldsymbol{\gamma})$ with respect to $\boldsymbol{\gamma}$. Following Zellner and Highfield (1988), Ormoneit

and White (1999), and Miller and Liu (2002), the set of necessary conditions are simply the moment conditions (14)–(16), and Newton–Raphson–type algorithms may be used to determine the solution values $\hat{\gamma}$ from these equations.

Although this computational procedure works in most entropy estimation problems, it is generally infeasible in stochastic frontier applications because the outcomes of u are not directly observed. However, we can follow the standard computational approach used in the stochastic frontier literature. First, u and v are assumed to be independent, and the joint density function may be formed by multiplication

$$f(u, v) = f(u) \times f(v) = \frac{u^{-\gamma_3} \exp\left(-u\gamma_1 - u^2\gamma_2 - \frac{v^2}{2\sigma_v^2}\right)}{\sigma_v \sqrt{2\pi} J_u(\gamma)} \quad (23)$$

Second, this outcome is restated to represent the joint density function of the composite error term $\varepsilon = v - u$ and u

$$f(u, \varepsilon) = \frac{u^{-\gamma_3} \exp\left(-u\gamma_1 - u^2\gamma_2 - \frac{(\varepsilon+u)^2}{2\sigma_v^2}\right)}{\sigma_v \sqrt{2\pi} J_u(\gamma)} \quad (24)$$

Third, the marginal probability density function for the composite error term ε may be derived by integration

$$\begin{aligned} f(\varepsilon) &= \int_0^\infty f(u, \varepsilon) du \\ &= \frac{\exp\left(-\frac{\varepsilon^2}{2\sigma_v^2}\right)}{\sigma_v \sqrt{2\pi} J_u(\gamma)} \int_0^\infty u^{-\gamma_3} \exp\left(-u\left(\gamma_1 + \frac{\varepsilon}{\sigma_v^2}\right) - u^2\left(\gamma_2 + \frac{1}{2\sigma_v^2}\right)\right) du \end{aligned} \quad (25)$$

The resulting integral also takes the form examined by Beckers and Hammond (1987), and the closed–form expression is denoted $J_\varepsilon(\varepsilon, \gamma, \sigma_v^2)$. Thus, the marginal density function for ε is

$$f(\varepsilon) = \frac{J_\varepsilon(\varepsilon, \gamma, \sigma_v^2) \exp\left(-\frac{\varepsilon^2}{2\sigma_v^2}\right)}{\sigma_v \sqrt{2\pi} J_u(\gamma)} \quad (26)$$

Following the traditional maximum likelihood estimation procedures for stochastic frontier models, the stochastic frontier model (1) is transformed under the natural logarithm, and the composite error $\varepsilon = \ln(y) - \ln(g(\mathbf{x}, \boldsymbol{\beta}))$ is substituted into the marginal probability density function (26) to form the log–likelihood function for an independent sample of n firms

$$\ell(\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma_v^2) \propto \sum_{i=1}^n \ln(J_\varepsilon(\varepsilon_i, \boldsymbol{\gamma}, \sigma_v^2)) - \frac{1}{2\sigma_v^2} \sum_{i=1}^n \varepsilon_i^2 - \frac{n}{2} \ln(\sigma_v^2) - n \ln(J_u(\boldsymbol{\gamma})) \quad (27)$$

The maximum likelihood estimation problem must be numerically solved, and the solution algorithm and estimated asymptotic covariance matrix may be based on the analytical gradients provided by Beckers and Hammond (1987). Conceptually, the

model parameters may also be estimated by the method of moments if the transformed stochastic frontier model is linear in the parameters

$$\ln(y) = (\beta_0 - E[u]) + \mathbf{x}\boldsymbol{\beta} + v - (u - E[u]) \quad (28)$$

and $\boldsymbol{\beta}$ may be consistently estimated by ordinary least squares (e.g., $g(\mathbf{x})$ is a Cobb–Douglas or translog model). Then, the sample moments of the composite residuals $\hat{\varepsilon}$ may be used to form a set of estimating equations for γ and σ_v^2 , and the model parameters may be estimated by the method of moments. The moment–based estimator is potentially consistent but inefficient relative to the maximum likelihood estimator. However, closed–form expressions for the moments of ε under the general stochastic frontier model (26) are not presently known, and the method of moments estimator is only feasible under a simulated moments approach to the estimation problem.

The maximum likelihood estimates may be used to conduct classical hypothesis tests in the usual way. For example, behavioral hypotheses related to elements of $\boldsymbol{\beta}$ in the regression function $g(\mathbf{x}, \boldsymbol{\beta})$ may be conducted under the Wald, Lagrange multiplier, or likelihood ratio criteria. More importantly, the log–likelihood function (27) may be used to compare the sample evidence regarding the parametric frontier models (e.g., exponential or half–normal) relative to the full set of traditional candidate models and the fitted version of the general frontier model. In particular, the parameters of the candidate models may be estimated by maximum likelihood or the method of moments and used to compute the log–likelihood value under this restriction. Then, the likelihood ratio criterion may be used to test the significance of the restrictions required to form one of the parametric frontier models relative to the fitted general frontier model. Alternatively, the log–likelihood values for each of the parametric frontier models may be used as a simple goodness–of–fit statistic to select the “best fitting” parametric frontier model without actually estimating the general form of the stochastic frontier model.

3.3 Illustrative example

To demonstrate the proposed method, we consider a data set originally studied by Zellner and Revankar (1970), who use $n = 25$ cross–sectional observations to estimate a Cobb–Douglas model of the value added by transportation equipment firms given capital and labor inputs. Several authors in this literature have used this data to compare the properties of competing stochastic frontier models, and the common conclusion is that the fitted exponential and half–normal models cannot be rejected because both provide very similar results. However, we can use the proposed techniques to conduct LM tests of these two alternative models relative to the fitted general frontier model.

For the exponential model, the null hypothesis is $H_0 : \gamma_2 = \gamma_3 = 0$ (i.e., two degrees of freedom), and the observed LM statistic is 113.36 (p-value is 0.00). Under the half-normal model, the null hypothesis is $H_0 : \gamma_3 = 0$ (i.e., one degree of freedom), and the observed LM statistic is 38.72 (p-value is 0.00). Thus, we strongly reject the exponential and half-normal models in favor of the general frontier model.

4 Measures of Technical Efficiency

4.1 Firm-specific technical efficiency

The estimated technical efficiency score for a given firm is based on the conditional probability density function

$$f(u | \varepsilon) = \frac{f(u, \varepsilon)}{f(\varepsilon)} = \frac{u^{-\gamma_3} \exp\left(-u\left(\gamma_1 + \frac{\varepsilon}{\sigma_v^2}\right) - u^2\left(\gamma_2 + \frac{1}{2\sigma_v^2}\right)\right)}{J_\varepsilon(\varepsilon, \boldsymbol{\gamma}, \sigma_v^2)} \quad (29)$$

with expected value

$$E[u | \varepsilon] = \frac{J_\varepsilon(\varepsilon, \boldsymbol{\gamma}^*, \sigma_v^2)}{J_\varepsilon(\varepsilon, \boldsymbol{\gamma}, \sigma_v^2)} \quad (30)$$

where $\boldsymbol{\gamma}^* = (\gamma_1, \gamma_2, \gamma_3 - 1)'$. One commonly used estimator of the technical efficiency score for a firm is then $\exp(E[u | \varepsilon])$. Alternatively, the technical efficiency score may be computed as

$$E[\exp(-u) | \varepsilon] = \frac{J_\varepsilon(\varepsilon, \boldsymbol{\gamma}^{**}, \sigma_v^2)}{J_\varepsilon(\varepsilon, \boldsymbol{\gamma}, \sigma_v^2)} \quad (31)$$

where $\boldsymbol{\gamma}^{**} = (\gamma_1 + 1, \gamma_2, \gamma_3)'$. It is important to note that these estimators are not consistent if the model is estimated from cross-sectional data. As with the traditional frontier models, consistent estimation of the firm-specific technical efficiency scores requires panel data.

4.2 Directed divergence and mean technical efficiency

Following the discussion from Section 2, the information theoretic concepts may be combined with Shephard's output distance function (1953, 1970) to develop measures of mean technical efficiency based on the distance between the regression functions $g(\mathbf{x})$ and $h(\mathbf{x})$. Shephard's output distance function measures the relationship between production outputs and the conditional mean (regression) function and is a useful tool for characterizing technical efficiency in production. The fundamental properties of the output distance function are outlined in Appendix A. Further, following the discussion in Section 2, a density function representation of the output distance function may be

developed for the single output case. In particular, let

$$\psi_g(y, \mathbf{x}) = D_0(y, \mathbf{x}) f(y, \mathbf{x}) \quad (32)$$

where $f(y, \mathbf{x})$ is the joint density function of y and \mathbf{x} . Given that $D_0(y, \mathbf{x}) = y/g(\mathbf{x})$ from (A.2),

$$\int \int \psi_g(y, \mathbf{x}) dy d\mathbf{x} = \int \frac{h(\mathbf{x}) f(\mathbf{x})}{g(\mathbf{x})} d\mathbf{x} \quad (33)$$

If $u = 0$ and there are no technical inefficiencies, $\psi_g(y, \mathbf{x})$ is a proper density function and integrates to one. Otherwise, $\psi_g(y, \mathbf{x})$ is improper and $\int \int \psi_g(y, \mathbf{x}) dy d\mathbf{x} < 1$, but $\psi_g(y, \mathbf{x})$ may be rescaled to represent a proper density function. A related density function that is always proper is

$$\psi_h(y, \mathbf{x}) = \frac{y f(y, \mathbf{x})}{h(\mathbf{x})} \quad (34)$$

which is based on the regression function, $h(\mathbf{x}) \equiv E[y | \mathbf{x}]$.

An information theoretic measure of mean technical efficiency can be derived as the information distance between $\psi_g(y, \mathbf{x})$ and $\psi_h(y, \mathbf{x})$. For continuous density functions $f_1(x)$ and $f_2(x)$, the Kullback–Leibler cross–entropy functional

$$I(f_1, f_2) = \int f_1(x) \ln \left(\frac{f_1(x)}{f_2(x)} \right) dx \quad (35)$$

measures the information content of $f_1(x)$ relative to the information expressed in $f_2(x)$. The support of x under f_1 must be a subset of the support of x under f_2 (i.e., f_2 is the dominating measure), and $I(f_1, f_2) = 0$ if $f_1(x) = f_2(x)$ (almost everywhere). The Kullback–Leibler (cross–entropy) functional is commonly known as a directed divergence measure because it is not symmetric, $I(f_1, f_2) \neq I(f_2, f_1)$, and is not a true distance function or measure. However, $I(f_1, f_2)$ may yet serve as a useful information criterion if we fix the reference distribution f_2 and consider the directed divergence of alternative f_1 candidates from this particular reference point.

Accordingly, the lack of symmetry in the Kullback–Leibler functional is not important for present purposes because the objective here is to measure the information in $\psi_h(y, \mathbf{x})$ relative to a reference distribution that represents the technically efficient case, $\psi_g(y, \mathbf{x})$. The cross–entropy between the density functions based on $D_0(y, \mathbf{x})$ is

$$I(\psi_h, \psi_g) = \int \int \psi_h(y, \mathbf{x}) \ln \left(\frac{\psi_h(y, \mathbf{x})}{\psi_g(y, \mathbf{x})} \right) dy d\mathbf{x} = \int f(\mathbf{x}) \ln \left(\frac{g(\mathbf{x})}{h(\mathbf{x})} \right) d\mathbf{x} \quad (36)$$

If firms are technically efficient, $g(\mathbf{x}) = h(\mathbf{x})$ (almost everywhere) and $I(\psi_h, \psi_g) = 0$. Otherwise, $g(\mathbf{x}) \geq h(\mathbf{x})$ (with strict inequality for some \mathbf{x}) such that $I(\psi_h, \psi_g) > 0$ in the presence of technical inefficiency. Thus, the Kullback–Leibler cross–entropy or

directed divergence functional provides an information theoretic criterion for evaluating the degree of mean technical efficiency in the production system.

Finally, we can show that the directed divergence measure in (36) may be simplified under the stochastic frontier model with multiplicative errors. By substitution of $h(\mathbf{x}) = g(\mathbf{x}) \text{E}[\exp(-u) | \mathbf{x}]$, we have

$$I(\psi_h, \psi_g) = - \int f(\mathbf{x}) \ln(\text{E}[\exp(-u) | \mathbf{x}]) \, d\mathbf{x} \quad (37)$$

Further, if the conditional mean $\text{E}[\exp(-u) | \mathbf{x}]$ is not a function of \mathbf{x} (as is commonly assumed), the directed divergence measure reduces to $I(\psi_h, \psi_g) = -\ln(\text{E}[\exp(-u)])$, which is a monotonic transformation of the mean technical efficiency stated above. For example, the mean technical efficiency measure for the half-normal frontier model $u \sim \text{N}^+(0, \sigma_u^2)$

$$\text{E}[\exp(-u)] = 2(1 - \Phi(\sigma_u)) \exp\left(\frac{\sigma_u^2}{2}\right) \quad (38)$$

may be stated in terms of the directed divergence measure as

$$I(\psi_h, \psi_g) = -\ln(2(1 - \Phi(\sigma_u))) - \frac{\sigma_u^2}{2} \geq 0 \quad (39)$$

Second, the mean technical efficiency measure for the exponential frontier model $u \sim \text{Exp}(\sigma_u)$ is

$$\text{E}[\exp(-u)] = (1 + \sigma_u)^{-1} \quad (40)$$

with directed divergence measure

$$I(\psi_h, \psi_g) = -\ln\left((1 + \sigma_u)^{-1}\right) = \ln(1 + \sigma_u) \geq 0 \quad (41)$$

In both cases, we have $I(\psi_h, \psi_g) = 0$ if production is technically efficient and $\sigma_u = 0$. Otherwise, $I(\psi_h, \psi_g) \rightarrow \infty$ as $\sigma_u \rightarrow \infty$ under technically inefficient production.

4.3 Remarks on formal tests of technical efficiency

Given estimates of the regression function $h(\mathbf{x})$ and the production frontier $g(\mathbf{x})$ based on a sample of n observations, formal tests of the technical efficiency hypothesis may be conducted in a variety of ways. However, for tests based on the estimated model parameters, we may have to take special care when conducting the tests if the null values of the model parameters lie on the boundary of the parameter space. For example, the null hypothesis associated with technical efficiency in the half-normal and exponential noise models is $H_0 : \sigma_u = 0$, and the usual asymptotic test statistics will not have the standard limiting distributions (Andrews, 2001).

5 Profile likelihood estimation

The normal model for v may be motivated as a maximum entropy distribution (as previously noted), but this model component is also subject to potential specification errors. To avoid potential limitations of the normality assumption, a profile likelihood estimation procedure based on semiparametric models of the production frontier $g(\mathbf{x})$ and the sampling error component (v) is proposed. One key advantage of the profile likelihood approach is that it is robust to model specification issues and provides a unified basis for model selection and other diagnostic tests. The extensions presented in this section follow from the original work reported by Miller (2002).

Assuming u and v are independent and are mutually uncorrelated with the elements of \mathbf{x} , the model parameters of stochastic frontier models are typically estimated from the logarithmic transformation of (1)

$$z = \ln(y) - \ln(g(\mathbf{x}, \boldsymbol{\beta})) = v - u \quad (42)$$

where $z \equiv \varepsilon$ for the “true” value of $\boldsymbol{\beta}$. Under technically efficient production, the method of moments (i.e., nonlinear least squares) estimator of $\boldsymbol{\beta}$ may be derived from the following set of moment equations

$$\mathbf{E} [\mathbf{G}'\mathbf{z} \mid \mathbf{x}] = \mathbf{0} \quad (43)$$

where $\mathbf{G} \equiv \nabla_{\boldsymbol{\beta}} g(\mathbf{x}, \boldsymbol{\beta})$ is the $(n \times k)$ Jacobian matrix of $g(\mathbf{x}, \boldsymbol{\beta})$. However, (43) does not hold with equality in the presence of technical inefficiencies because

$$\mathbf{E}[z \mid \mathbf{x}] = -\mathbf{E}[u \mid \mathbf{x}] \quad (44)$$

In this case, researchers commonly use fully parametric model specifications and estimate the unknown parameters by maximum likelihood.

In contrast to the approach taken in Section 3, the semiparametric estimation method proposed by Miller (2002) incorporates a fully parametric specification for u and builds a flexible model of v . As noted above, the traditional stochastic frontier models for u (e.g., exponential or half-normal) are functions of the parameter σ_u^2 , and the model of the composite error term ε also depends on the regression parameters $\boldsymbol{\beta} \in \mathbb{R}^k$ and the variance of the other error term, σ_v^2 . The resulting set of $(k+2)$ model parameters may be expressed in terms of the following $(k+2)$ estimating equations

$$\mathbf{E} [\mathbf{G}'(z + \mathbf{E}[u]) \mid \mathbf{x}] = \mathbf{0} \quad (45)$$

$$\mathbf{E}[z + \mathbf{E}[u] \mid \mathbf{x}] = \mathbf{0} \quad (46)$$

$$\mathbf{E} \left[(z + \mathbf{E}[u])^2 - \text{var}(u) - \sigma_v^2 \mid \mathbf{x} \right] = \mathbf{0} \quad (47)$$

when stated in homogeneous form. To derive a class of moment-based estimators of the model parameters, Miller (2002) forms the empirical analog functions from (45)–(47)

$$\sum_{i=1}^n \pi_i \mathbf{G}'_i [\ln(y_i) - g(\mathbf{x}_i, \boldsymbol{\beta}) + E[u]] = \mathbf{0} \quad (48)$$

$$\sum_{i=1}^n \pi_i [\ln(y_i) - g(\mathbf{x}_i, \boldsymbol{\beta}) + E[u]] = 0 \quad (49)$$

$$\sum_{i=1}^n \pi_i \left\{ [\ln(y_i) - g(\mathbf{x}_i, \boldsymbol{\beta}) + E[u]]^2 - \text{var}(u) - \sigma_v^2 \right\} = 0 \quad (50)$$

where π_i is the empirical weight assigned to observation i . The empirical weights must satisfy non-negativity ($\pi_i \geq 0$) and the additivity property, $\sum_{i=1}^n \pi_i = 1$. The moment-based estimation problem is solved by simultaneously selecting the empirical weights and the parameters in order to satisfy the set of $(k+2)$ equations.

Owen (1991) and Qin and Lawless (1994) describe the empirical likelihood procedure for solving estimation problems of this type, and Imbens (1997) and Kitamura and Stutzer (1997) propose an entropy-based alternative based on Shannon's entropy functional

$$H(\boldsymbol{\pi}) = - \sum_{i=1}^n \pi_i \ln(\pi_i) \quad (51)$$

for discrete random variables. The objective of the entropy-based approach is to derive a profile likelihood function for the unknown model parameters by maximizing (51) subject to (48)–(50) and the additivity constraint, $\sum_{i=1}^n \pi_i = 1$. Under Jaynes' method of maximum entropy, the maximum entropy empirical weights are

$$\hat{\pi}_i = \frac{\exp(-\hat{\boldsymbol{\alpha}}' \mathbf{G}'_i (z_i + E[u]) - \hat{\gamma}_1 (z_i + E[u]) - \hat{\gamma}_2 [(z_i + E[u])^2 - \text{var}(u) - \sigma_v^2])}{\Omega} \quad (52)$$

and $\Omega \equiv \sum_{i=1}^n \exp(\hat{\boldsymbol{\alpha}}' \mathbf{G}'_i (z_i + E[u]) - \hat{\gamma}_1 (z_i + E[u]) - \hat{\gamma}_2 [(z_i + E[u])^2 - \text{var}(u) - \sigma_v^2])$ is the normalizing constant. The empirical weights are closest to the uniform weights $\pi_i = n^{-1}$ while satisfying the required moment conditions. The elements $\hat{\boldsymbol{\alpha}}$, $\hat{\gamma}_1$, and $\hat{\gamma}_2$ are the optimal values of the Lagrange multipliers on constraints (48)–(50). By substitution of (52) back into (51), the profile likelihood function for the unknown parameters $(\boldsymbol{\beta}, \sigma_u, \sigma_v^2)$ and the Lagrange multipliers is formed. The profile likelihood function is optimized to jointly determine the stochastic frontier parameter estimates and the optimal Lagrange multipliers.

Under suitable regularity conditions, the entropy-based profile likelihood estimators of the stochastic frontier parameters are consistent, asymptotically normal, and asymptotically equivalent to empirical likelihood estimators. Consequently, the estimated asymptotic distributions of the entropy-based estimator may be used to conduct

asymptotic tests related to the model parameters. Further, Monte Carlo evidence suggests that the entropy-based approach to profile likelihood estimation may be more robust to specification errors in the moment conditions, which is a potentially important advantage of this approach over fully parametric likelihood estimation in stochastic frontier analysis. For more details on the computational properties of the entropy-based profile likelihood approach, please refer to Chapter 13 in Mittelhammer, Judge, and Miller (2000).

6 Summary and Conclusions

This paper considers the problem of specification, estimation, and testing of stochastic frontier models from an information theoretic perspective. Given the basic properties of the parametric stochastic frontier models commonly used in practice, the maximum entropy procedure developed by Jaynes is used to construct a general stochastic frontier model that nests all of the traditional parametric alternatives. Although the proposed model must be estimated with computationally intensive procedures, the required computational burden is comparable to the effort required to estimate the existing models. Further, the proposed stochastic frontier model provides a unified framework for classical tests of behavioral hypotheses and model goodness-of-fit because it nests the parametric models as special cases. The framework is also extended to develop information theoretic measures of mean technical efficiency and profile likelihood estimation procedures based on a set of relevant estimating equations. The resulting profile likelihood function may be used to derive consistent and asymptotically normal estimators of the unknown parameters of a stochastic frontier model, but it requires a particular distributional assumption regarding the frontier error component u . Efforts to form a fully semiparametric version of this model are currently underway.

Information theoretic methods have been used in the production economics literature for a long time and in some substantially different ways. For example, the early work presented by Georgescu-Roegen (1971) is more closely associated with the physical interpretation of entropy, which is fundamentally different than the concept of entropy developed in the statistical information theory literature. As well, Sengupta (1992) used the information theoretic approach to analyze frontier models, but his approach is more closely related to DEA than to stochastic frontier modeling. This paper shows that the information theoretic interpretation of density functions, regression functions, and distance functions provides a rich framework for building unifying links among the various segments of the production economics literature. For example, the tools developed in this paper may be extended to derive information theoretic

measures of technical efficiency and other behavioral outcomes associated with input distance functions (Chambers, Chung, and Fare, 1996) and directional distance functions (Fare and Grosskopf, 2000) plus the associated dual functions. Second, the information theoretic tools may be used to develop more comprehensive models of multi-output production systems (Fare and Primont, 1990). Finally, the present work focuses on cross-sectional production outcomes, and the proposed methods may be readily extended to represent panel data production outcomes under time-invariant or time-varying technical efficiency.

References

- Aigner, D. and D.-S. Chu**, “On Estimating the Industry Production Function,” *American Economic Review*, 1968, *58*, 826–39.
- , **C. A. K. Lovell, and P. Schmidt**, “Formulation and Estimation of Stochastic Frontier Production Function Models,” *Journal of Econometrics*, 1977, *6*, 21–37.
- Andrews, D. W. K.**, “Testing when a Parameter is on the Boundary of a Maintained Hypothesis,” *Econometrica*, 2001, *69*, 683–734.
- Beckers, D. and C. Hammond**, “A Tractable Likelihood Function for the Normal–Gamma Stochastic Frontier Model,” *Economics Letters*, 1987, *24*, 33–38.
- Chambers, R. G.**, *Applied Production Analysis*, New York: Cambridge University Press, 1988.
- , **Y. Chung, and R. Fare**, “Benefit Distance Functions,” *Journal of Economic Theory*, 1996, *70*, 407–19.
- Coelli, T.**, “Estimators and Hypothesis Tests for a Stochastic Frontier Function: A Monte Carlo Analysis,” *Journal of Productivity Analysis*, 1995, *6*, 247–68.
- Cornes, R.**, *Duality and Modern Economics*, New York: Cambridge University Press, 1992.
- Fare, R.**, *Fundamentals of Production Theory*, Heidelberg: Springer–Verlag, 1988.
- and **D. Primont**, “A Distance Function Approach to Multioutput Technologies,” *Southern Economic Journal*, 1990, *56*, 879–91.
- and **S. Grosskopf**, “Theory and Application of Directional Distance Functions,” *Journal of Productivity Analysis*, 2000, *13*, 93–103.
- Fuss, M. and D. McFadden**, *Production Economics: A Dual Approach to Theory and Applications*, Amsterdam: North Holland, 1978.
- Georgescu-Roegen, N.**, *The Entropy Law and the Economic Process*, Cambridge, MA: Harvard University Press, 1971.
- Golan, A., G. Judge, and D. Miller**, *Maximum Entropy Econometrics: Robust Estimation with Limited Information*, Chichester, UK: John Wiley and Sons, 1996.
- Gong, B.-H. and R. Sickles**, “Finite Sample Evidence on the Performance of Stochastic Frontiers and Data Envelopment Analysis using Panel Data,” *Journal of Productivity Analysis*, 1989, *1*, 229–61.
- and —, “Finite Sample Evidence on the Performance of Stochastic Frontiers and Data Envelopment Analysis using Panel Data,” *Journal of Econometrics*, 1992, *51*, 259–84.

- Greene, W.**, “Maximum Likelihood Estimation of Econometric Frontier Functions,” *Journal of Econometrics*, 1980, *13*, 27–56.
- , “On the Estimation of a Flexible Frontier Production Model,” *Journal of Econometrics*, 1980, *13*, 101–15.
- , “A Gamma–Distributed Stochastic Frontier Model,” *J. Econometrics*, 1990, *46*, 141–64.
- Imbens, G.**, “One–Step Estimators for Over–Identified Generalized Method of Moment Models,” *Review of Economic Studies*, 1997, *64*, 359–83.
- Jaynes, E. T.**, “Information Theory and Statistical Mechanics, I,” *Physical Review*, 1957, *106*, 620–30.
- , “Information Theory and Statistical Mechanics, II,” *Physical Review*, 1957, *108*, 171–90.
- Kitamura, Y. and M. Stutzer**, “An Information Theoretic Alternative to Generalized Method of Moments Estimation,” *Econometrica*, 1997, *65*, 861–74.
- Kullback, S.**, *Information Theory and Statistics*, New York: John Wiley and Sons, 1959.
- Kuumbhakar, S. and C. A. K. Lovell**, *Stochastic Frontier Analysis*, New York: Cambridge University Press, 2000.
- Lee, L.-F. and W. Tyler**, “The Stochastic Frontier Production Function and Average Efficiency: An Empirical Analysis,” *Journal of Econometrics*, 1978, *7*, 385–89.
- Massoumi, E.**, “A Compendium to Information Theory in Economics and Econometrics,” *Econometric Reviews*, 1993, *12*, 137–81.
- Meeusen, W. and J. van den Broeck**, “Efficiency Estimation from Cobb–Douglas Production Functions with Composed Error,” *International Economic Review*, 1977, *18*, 435–44.
- Miller, D.**, “Entropy–Based Methods of Modeling Stochastic Production Efficiency,” *American Journal of Agricultural Economics*, 2002, *84*, 1264–70.
- and **W.-H. Liu**, “On the Recovery of Joint Distributions from Limited Information,” *Journal of Econometrics*, 2002, *107*, 259–74.
- Mittelhammer, R., G. Judge, and D. Miller**, *Econometric Foundations*, New York, NY: Cambridge University Press, 2000.
- Ormoneit, D. and H. White**, “An Efficient Algorithm to Compute Maximum Entropy Densities,” *Econometric Reviews*, 1999, *18*, 127–40.

- Owen, A.**, “Empirical Likelihood for Linear Models,” *Annals of Statistics*, 1991, *19*, 1725–47.
- Qin, J. and J. Lawless**, “Empirical Likelihood and General Estimating Equations,” *Annals of Statistics*, 1994, *22*, 300–25.
- Ryu, H.**, “Maximum Entropy Estimation of Density and Regression Functions,” *Journal of Econometrics*, 1993, *56*, 397–440.
- Schmidt, P. and T.-F. Lin**, “Simple Tests of Alternative Specifications in Stochastic Frontier Models,” *Journal of Econometrics*, 1984, *24*, 349–61.
- Sengupta, J.**, “The Maximum Entropy Approach in Production Frontier Estimation,” *Mathematical Social Sciences*, 1992, *25*, 41–57.
- Shephard, R.**, *Cost and Production Functions*, Princeton, NJ: Princeton Univ. Press, 1953.
- , *The Theory of Cost and Production Functions*, Princeton, NJ: Princeton U. Press, 1970.
- Sickles, R., D. Good, and L. Getachew**, “Specification of Distance Functions Using Semi- and Nonparametric Methods with an Application to the Dynamic Performance of Eastern and Western European Air Carriers,” *Journal of Productivity Analysis*, 2002, *17*, 133–55.
- Stevenson, R.**, “Likelihood Functions for Generalized Stochastic Frontier Estimation,” *Journal of Econometrics*, 1980, *13*, 57–66.
- Theil, H. and D. Fiebig**, *Exploiting Continuity: Maximum Entropy Estimation of Continuous Distributions*, Cambridge, MA: Ballinger Publishing, 1984.
- Zellner, A. and N. Revankar**, “Generalized Production Functions,” *Review of Economic Studies*, 1970, *37*, 241–50.
- and **R. Highfield**, “Calculation of Maximum Entropy Distributions and Approximation of Marginal Posterior Distributions,” *Journal of Econometrics*, 1988, *37*, 195–209.

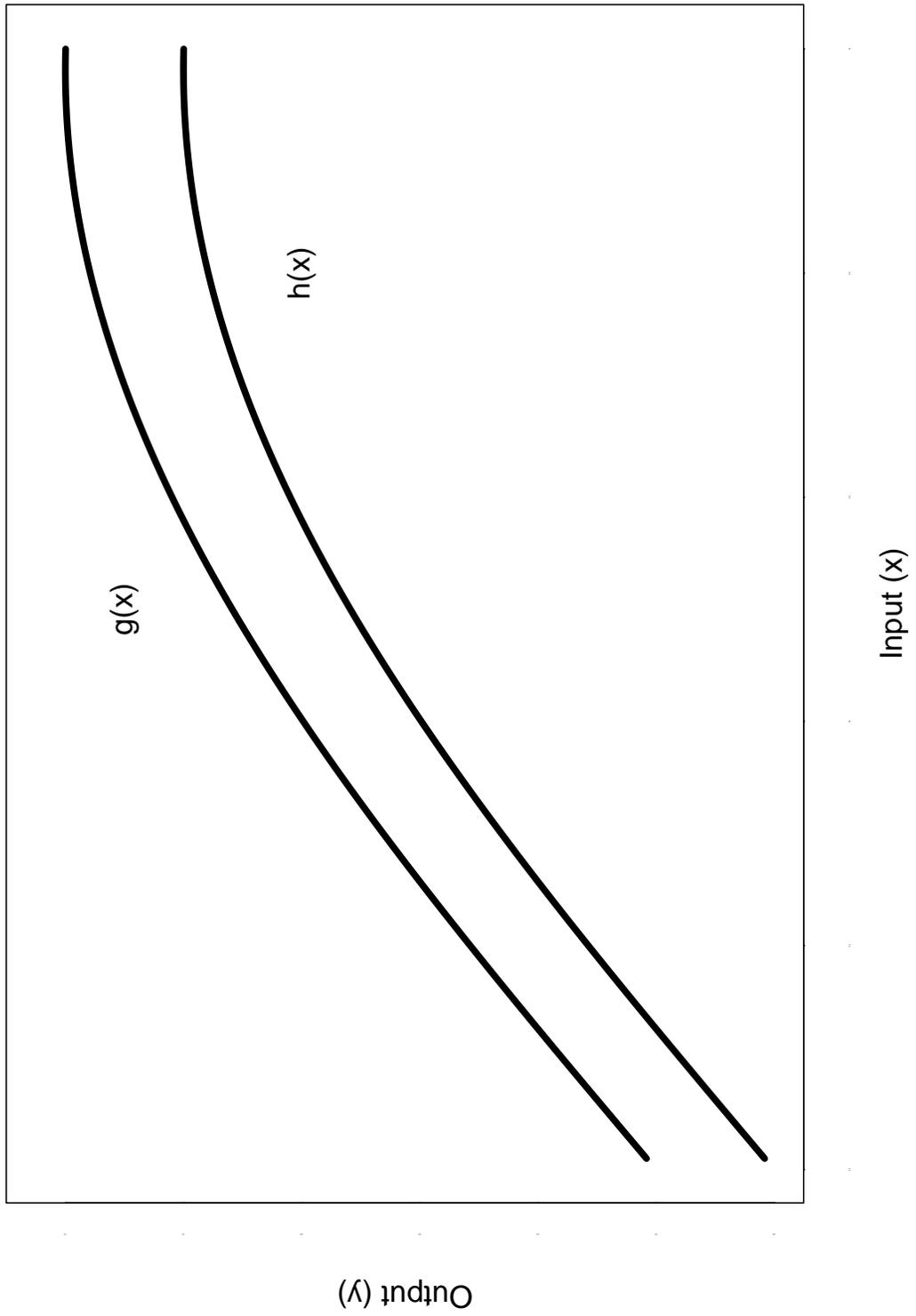


Figure 1: Production Frontier (single-input, single-output case)

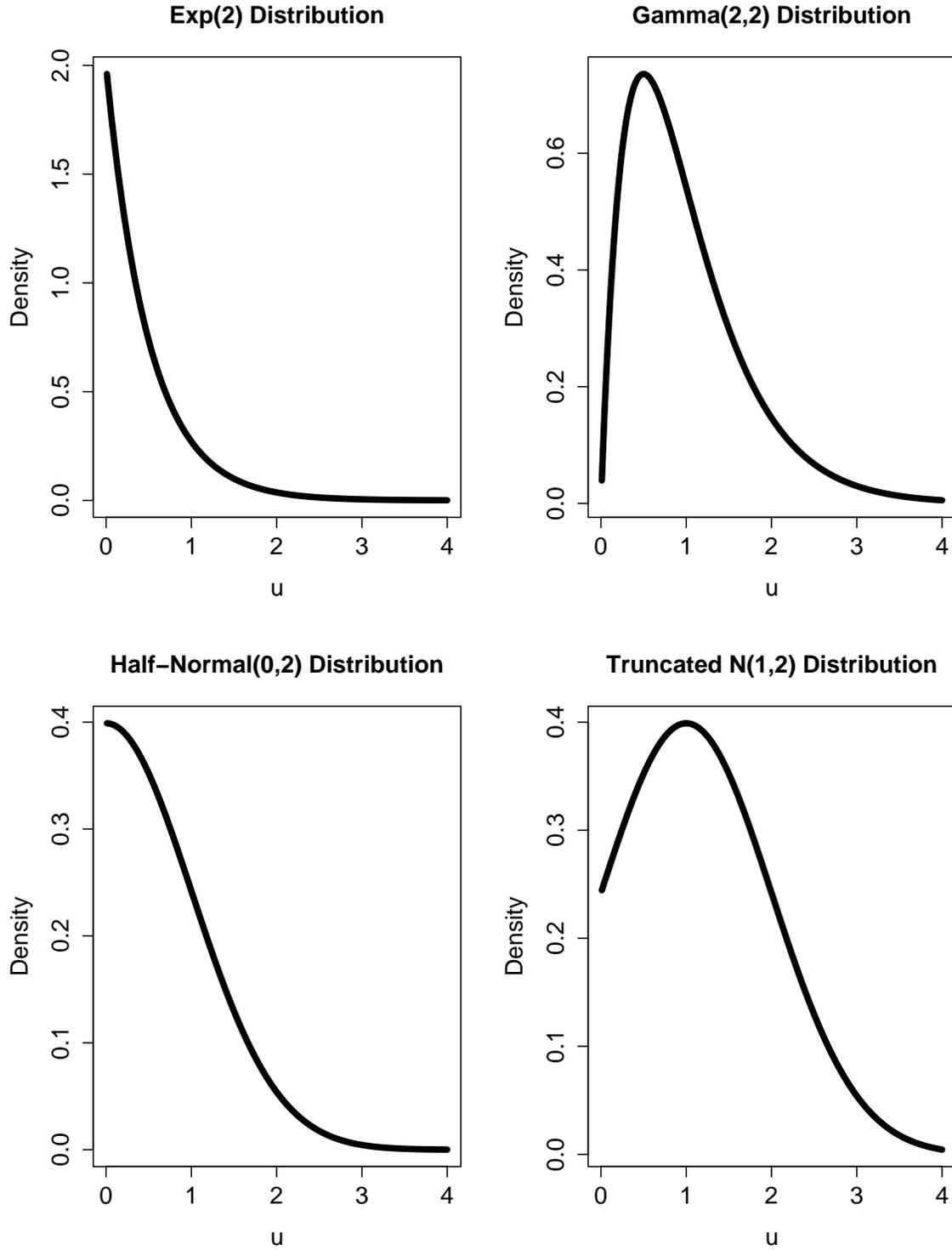


Figure 2: Exponential, Gamma, Half-Normal, and Truncated Normal PDF's

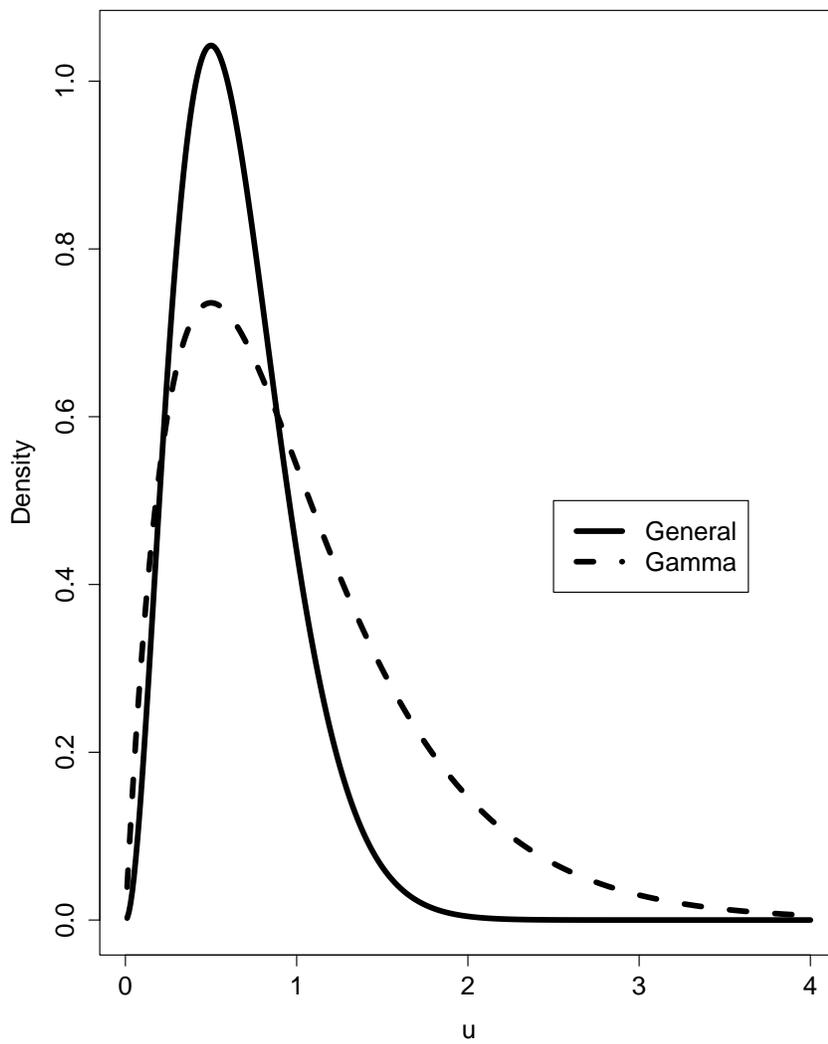


Figure 3: Gamma and General Frontier Models (same mode)

Table 1: Moments of the Candidate Frontier Models

Model	Mean	Variance
Exponential(σ_u)	σ_u	σ_u^2
Half-normal($0, \sigma_u^2$)	$\sigma_u \sqrt{\frac{2}{\pi}}$	$\sigma_u^2 \left(\frac{\pi-2}{\pi}\right)$
Truncated normal(μ, σ_u^2)	$\frac{\mu a}{2} + \frac{\sigma_u a}{\sqrt{2\pi}} \exp\left(-\frac{\mu^2}{2\sigma_u^2}\right)$	$\mu^2 \frac{a}{2} \left(1 - \frac{a}{2}\right) + \frac{a}{2} \left(\frac{\pi-2}{\pi}\right) \sigma_u^2$
Gamma(m, σ_u)	$(m+1)\sigma_u$	$(m+1)\sigma_u^2$

For the truncated normal model, $a = [\Phi(\mu/\sigma_u)]^{-1}$

Table 2: Maximum Entropy Constraints and Lagrange Multipliers

Model	First Constraint		Second Constraint	
	Moment	Lagrange multiplier	Moment	Lagrange multiplier
Exponential(σ_u)	$E[u]$	σ_u^{-1}		
Half-normal($0, \sigma_u^2$)	$\text{var}(u)$	$(2\sigma_u^2)^{-1}$		
Truncated normal(μ, σ_u^2)	$E[u]$	$\mu\sigma_u^{-2}$	$\text{var}(u)$	$(2\sigma_u^2)^{-1}$
Gamma(m, σ_u)	$E[u]$	σ_u^{-1}	$E[\ln(u)]$	$-m$

Table 3: Other Special Cases of the General Frontier Model

Model	γ_1	γ_2	γ_3
Chi(a)	0	0.5	$1 - a$
Erlang(a, b)	b	0	$1 - a$
Exponential power($a, b = 2$)	0	a^{-2}	0
Nakagami(a, b)	0	a/b	$1 - 2a$
Rayleigh(a)	0	$(2a^2)^{-1}$	-1
Weibull($a = 1, b$)	b^{-1}	0	0
Weibull($a = 2, b$)	0	b^{-2}	-1

Appendix A: Properties of Output Distance Functions

Shephard (1953, 1970) introduced the distance function approach to production analysis, and the general theory and properties have been developed by Fuss and McFadden (1978), Fare (1988), Cornes (1992), and Kumbhakar and Lovell (2000). To begin, suppose the single output y is generated from k inputs $\mathbf{x} = (x_1, \dots, x_k)$. Let $P(\mathbf{x})$ denote the feasible set of output values that can be produced from input vector, \mathbf{x} , and the assumed properties of the output set are

- $P(\mathbf{0}) = \{0\}$
- $P(\mathbf{x})$ is a closed set
- $P(\mathbf{x})$ is bounded for $\mathbf{x} \in \mathbb{R}_+^k$
- $P(\lambda\mathbf{x}) \supseteq P(\mathbf{x})$ for $\lambda \geq 1$
- $y \in P(\mathbf{x})$ implies $\lambda y \in P(\mathbf{x})$ for $\lambda \in [0, 1]$

Accordingly, the *output distance function* is defined as

$$D_0(y, \mathbf{x}) = \min \{ \theta : y/\theta \in P(\mathbf{x}) \} \quad (\text{A.1})$$

(see Definition 2.11 in Kumbhakar and Lovell, 2000). The associated properties of the output distance function are

- $D_0(0, \mathbf{x}) = 0$ and $D_0(y, \mathbf{0}) = \infty$
- $D_0(y, \mathbf{x})$ is lower-semicontinuous
- $D_0(\lambda y, \mathbf{x}) = \lambda D_0(y, \mathbf{x})$ for $\lambda > 0$
- $D_0(y, \lambda\mathbf{x}) \leq D_0(y, \mathbf{x})$ for $\lambda \geq 1$
- $D_0(\lambda y, \mathbf{x}) \leq D_0(y, \mathbf{x})$ for $\lambda \in [0, 1]$

The output set may be equivalently defined in terms of the output distance function $D_0(y, \mathbf{x})$ as $P(\mathbf{x}) = \{y : D_0(y, \mathbf{x}) \leq 1\}$.

The output distance function approach is very useful for measurement of technical efficiency. In particular, the (y, \mathbf{x}) outcomes represent technically efficient production if $D_0(y, \mathbf{x}) = 1$, and production is technically inefficient if $D_0(y, \mathbf{x}) < 1$. In the case of a single output, the output distance function reduces to

$$D_0(y, \mathbf{x}) = \frac{y}{g(\mathbf{x})} \quad (\text{A.2})$$

where $g(\mathbf{x})$ denotes the deterministic (conditional on \mathbf{x}) production frontier

$$g(\mathbf{x}) = \max \{y : y \in P(\mathbf{x})\} \quad (\text{A.3})$$

(see Definition 2.9 in Kumbhakar and Lovell, 2000). Output levels that achieve the production frontier, $y = g(\mathbf{x})$, are technically efficient such that $D_0(y, \mathbf{x}) = 1$, and $D_0(y, \mathbf{x}) < 1$ in the technically inefficient case. Accordingly, the stochastic distance function model takes the form

$$1 = D_0(y, \mathbf{x}) \exp(u - v) \tag{A.4}$$

Following the notation established in Section 1, v represents measurement error in the output level and u is a firm-specific random effect that represents the firm's technical inefficiency. As such, v is assumed to be a mean-zero error term that is symmetrically distributed about zero, and u is an error term supported on the positive real line. In the single output case, the distance function (A.4) implies that the stochastic frontier model takes the form stated in (1).