Journal of Econometrics 6 (1977) 21–37. © North-Holland Publishing Company

FORMULATION AND ESTIMATION OF STOCHASTIC FRONTIER PRODUCTION FUNCTION MODELS*

Dennis AIGNER

University of Wisconsin, Madison, WI 53706, U.S.A.
University of Southern California, Los Angeles, CA 90007, U.S.A.

C.A. Knox LOVELL

University of North Carolina, Chapel Hill, NC 27514, U.S.A.

Peter SCHMIDT

University of North Carolina, Chapel Hill, NC 27514, U.S.A.

Received October 1975, final version received October 1976

Previous studies of the so-called frontier production function have not utilized an adequate characterization of the disturbance term for such a model. In this paper we provide an appropriate specification, by defining the disturbance term as the sum of symmetric normal and (negative) half-normal random variables. Various aspects of maximum-likelihood estimation for the coefficients of a production function with an additive disturbance term of this sort are then considered.

1. Introduction

The theoretical definition of a production function expressing the *maximum* amount of output obtainable from given input bundles with fixed technology has been accepted for many decades. And for almost as long, econometricians have been estimating *average* production functions. It has only been since the pioneering work of Farrell (1957) that serious consideration has been given to the possibility of estimating so-called frontier production functions, in an effort to bridge the gap between theory and empirical work. For a variety of reasons these efforts have not been completely successful. In this paper we suggest a new approach to the estimation of frontier production functions. This involves the specification of the error term as being made up of two components, one normal and the other from a one-sided distribution. This approach enables us to overcome some of the major shortcomings of previous work in the area.

*This work was supported in part by NSF Grant GS-39995 (DJA). We are indebted to Takeshi Amemiya, Forrest Nelson, and Dale J. Poirier for helpful advice at various stages of the research, and to Chinbang Chung for research assistance. None of them should be held responsible for any errors that remain.

The outline of the paper is as follows. In section 2 we review previous approaches to the estimation of parametric frontier production functions, and in section 3 we propose a new approach. The statistical properties of our model are discussed in section 4. Some Monte Carlo experiments are reported in section 5, and empirical examples are given in section 6. Section 7 concludes.

2. Previous parametric frontier models

Previous work on the estimation of parametric frontier production functions, as characterized by the work of Aigner and Chu (1968), Afriat (1972) and Richmond (1974), begins by assuming a function giving maximum possible output as a function of certain inputs. For a given firm, say the ith, we write

$$y_i = f(\boldsymbol{x}_i; \boldsymbol{\beta}). \tag{1}$$

Here y_i is the maximum output obtainable from $\boldsymbol{x}_i$, a vector of (non-stochastic) inputs, and $\boldsymbol{\beta}$ is an unknown parameter vector to be estimated.

Aigner and Chu (1968) suggest the estimation of $\boldsymbol{\beta}$ by mathematical programming methods based on a cross-section of N firms within a given industry. Specifically, they suggest minimization of

$$\sum_{i=1}^{N} |y_i - f(\boldsymbol{x}_i; \boldsymbol{\beta})|,$$

subject to $y_i \leqq f(\boldsymbol{x}_i; \boldsymbol{\beta})$, which is a linear programming problem if $f(\boldsymbol{x}_i; \boldsymbol{\beta})$ is linear in $\boldsymbol{\beta}$. Alternatively, they suggest minimization of

$$\sum_{i=1}^{N} [y_i - f(\boldsymbol{x}_i; \boldsymbol{\beta})]^2,$$

subject to the same constraint, which is a quadratic programming problem if $f(\boldsymbol{x}_i; \boldsymbol{\beta})$ is linear.

Obviously, something magical has happened in moving from (1) to either of these 'estimation' methods: In order to characterize differences in output among firms with identical input vectors or to explain how a given firm's output lies below the 'frontier', $f(\boldsymbol{x}_i, \boldsymbol{\beta})$, a disturbance term has been implicitly assumed.

One problem with these approaches is extreme sensitivity to outliers. This has led to the development of so-called 'probabilistic' frontiers [Timmer (1971), Dugger (1974)] which are estimated by the same types of mathematical programming techniques discussed above, except that some specified proportion of the observations is allowed to lie above the frontier. The selection of this proportion is essentially arbitrary, lacking explicit economic or statistical justification. Another problem involves reconciling the observations above the frontier with the concept of the frontier as maximum possible output. Typically this is accomplished by appealing to measurement error in the extreme observations. However, it seems preferable to incorporate the possibility of measurement error, and of other unobservable shocks, in a less arbitrary fashion.

As they have been applied previously, therefore, the mathematical programming techniques do not lead to estimates with known statistical properties. In an attempt to give them a statistical basis, Schmidt (1976) explicitly added a one-sided disturbance to (1) above, which yields the model

$$y_i = f(\boldsymbol{x}_i; \boldsymbol{\beta}) + \varepsilon_i, \qquad i = 1, \ldots, N, \tag{2}$$

where $\varepsilon_i \leqq 0$. Given a distribution assumption for the disturbance term, the model can then be estimated by maximum-likelihood techniques. In particular, the assumption that $-\varepsilon_i$ has an exponential distribution leads to the linear programming technique, while the assumption that $-\varepsilon_i$ has a half-normal distribution leads to the quadratic programming technique. Therefore, Aigner and Chu's estimates can be viewed as maximum-likelihood estimates under particular error specifications.

Unfortunately, the observation that the model can be estimated by maximum-likelihood techniques, and that under appropriate assumptions linear and quadratic programming are maximum-likelihood techniques, is of little practical value. This is so because the usual 'regularity conditions' for the application of maximum likelihood are violated. In particular, since $y_i \leqq f(\boldsymbol{x}_i; \boldsymbol{\beta})$, the range of the random variable y depends on the parameters to be estimated. Therefore, the usual theorems cannot be invoked to determine the asymptotic distributions of parameter estimates. Under these circumstances it is not clear just how much we know about the frontier after having estimated it.

In another recent paper, Aigner, Amemiya, and Poirier (1976) construct a more reasonable error structure than a purely one-sided one. Specifically, they assume

$$\varepsilon_i = \begin{cases} \varepsilon_i^*/\sqrt{(1-\theta)}, & \text{if} \quad \varepsilon_i^* > 0, \quad i = 1, \ldots, N, \\ \varepsilon_i^*/\sqrt{\theta}, & \text{if} \quad \varepsilon_i^* \leqq 0, \end{cases} \tag{3}$$

where the errors, ε_i^*, are independent normally distributed random variables with zero means and variance σ^2 for $0 < \theta < 1$; otherwise, ε_i^* has either the negative or positive truncated normal distribution, when $\theta = 1$ or $\theta = 0$, respectively.

Their justification for this error specification is that firms are presumed to differ in their 'production' of y for a given set of values for the 'inputs' according to random variation in (1) their ability to utilize 'best practice' technology, a source of error that is one-sided ($\varepsilon_i \leqq 0$), and/or (2) an input quantity or measurement error in y, a symmetric error. The parameter θ is interpreted as the measure of 'relative variability' in these two error sources, its values circumscribing the 'full' frontier function ($\theta = 1$), the 'average' function ($\theta = \frac{1}{2}$), and intermediate cases of some interest.[1]

A primary contribution of this error structure to the literature is that it allows the placement of the fitted function to be estimated along with the usual parameters of interest through the parameter θ. Thus, the criticism levied at the average function by proponents of the frontier [e.g., Aigner and Chu (1968)] and criticisms that accompany strict use of the frontier or envelope function as the 'appropriate' industry production function [cf., Timmer (1971)] are ameliorated by this more accommodating specification.

Nevertheless, the interpretation of θ as a measure of the relative variability of error sources is only implicit in the Aigner, Amemiya, Poirier formulation A more direct approach is to specifically model the error process implied by the behavioral considerations mentioned above.

3. A stochastic frontier

We now return to the model as given in eq. (2), but under the error structure

$$\varepsilon_i = v_i + u_i, \qquad i = 1, \ldots, N. \tag{4}$$

The error component v_i represents the symmetric disturbance: the $\{v_i\}$ are assumed to be independently and identically distributed as $N(0, \sigma_v^2)$. The error component u_i is assumed to be distributed independently of v_i, and to satisfy $u_i \leqq 0$. We will be particularly concerned with the case in which u_i is derived from a $N(0, \sigma_u^2)$ distribution truncated above at zero. However, other one-sided distributions are tenable, and we will also briefly consider the case in which $-u_i$ has an exponential distribution.

This model collapses to a deterministic frontier model when $\sigma_v^2 = 0$, and it collapses to the Zellner, Kmenta, and Drèze (1966) stochastic production function model when $\sigma_u^2 = 0$. Note that $y_i \leqq f(\mathbf{x}_i; \boldsymbol{\beta}) + v_i$, so that the frontier itself is now clearly stochastic.

The economic logic behind this specification is that the production process is subject to two economically distinguishable random disturbances, with different characteristics. We believe that there is ample precedent in the literature for

[1]As $\theta \to 1$, the positive error component has a large variance (hence small influence in the likelihood function), and the negative error dominates. This gives rise to the 'full' frontier as the limiting case ($\theta = 1$). A similar interpretation follows for the case of $\theta \to 0$, although a behavioral explanation for this situation is lacking. When $\theta = \frac{1}{2}$, the likelihood function has the form of a mixture of two half-normals, each with equal influence.

such a view, although our interpretation is clearly new.[2] And from a practical standpoint, such a distinction greatly facilitates the estimation and interpretation of a frontier. The non-positive disturbance u_i reflects the fact that each firm's output must lie on or below its frontier $[f(\boldsymbol{x}_i; \boldsymbol{\beta})+v_i]$. Any such deviation is the result of factors under the firm's control, such as technical and economic inefficiency, the will and effort of the producer and his employees, and perhaps such factors as defective and damaged product. But the frontier itself can vary randomly across firms, or over time for the same firm. On this interpretation, the frontier is stochastic, with random disturbance $v_i \gtreqless 0$ being the result of favorable as well as unfavorable external events such as luck, climate, topography, and machine performance. Errors of observation and measurement on y constitute another source of $v_i \gtreqless 0$.

One interesting byproduct of this approach is that we can *estimate* the variances of v_i and u_i, so as to get evidence on their relative sizes. Another implication of this approach is that productive efficiency should, in principle, be measured by the ratio

$$y_i/[f(\boldsymbol{x}_i; \boldsymbol{\beta})+v_i], \tag{5}$$

rather than by the ratio

$$y_i/[f(\boldsymbol{x}_i; \boldsymbol{\beta}]. \tag{6}$$

This simply distinguishes productive inefficiency from other sources of disturbance that are beyond the firm's control. For example, the farmer whose crop is decimated by drought or storm is unlucky on our measure (5), but inefficient by the usual measure (6).[3]

Our discussion of estimation will be simplified somewhat if we consider a linear model. We therefore write, in obvious matrix form,

$$\boldsymbol{y} = X\boldsymbol{\beta}+\boldsymbol{\varepsilon}, \tag{7}$$

in place of (2), where now $\boldsymbol{\varepsilon} = \boldsymbol{v}+\boldsymbol{u}$.

[2]Marschak and Andrews (1944) suggest that the sum (v_i+u_i) reflects the 'technical efficiency' and the 'will, effort and luck' of a producer. Zellner, Kmenta, and Drèze (1966) suggest that it reflects 'factors such as weather, unpredictable variations in machine or labor performance, and so on', and they were perhaps the first to propose a stochastic production function, although they clearly did not have a frontier in mind. Other characterizations of the error term exist: Aigner and Chu (1968) explain it by reason of technical and economic inefficiencies, as well as by pure random shocks in the production process that might be due to careless handling and defective or damaged output. Timmer (1971) cites technical and economic inefficiency, as well as 'definitional and measurement problems in the variables'. And, agricultural economists frequently cite variation across farms in such environmental conditions as climate, topography and soil type as indicative of a random production function.

[3]As defined, (5) is a strange efficiency measure, since it is stochastic and depends on an unobservable, v_i. It is offered here only to support the argument.

4. Estimation of the stochastic frontier model

The distribution function of the sum of a symmetric normal random variable and a truncated normal random variable was apparently first derived by M.A. Weinstein (1964). The derivation of the density function of ε is straightforward, so we shall not include it here. The result is

$$f(\varepsilon) = \frac{2}{\sigma} f^*\left(\frac{\varepsilon}{\sigma}\right)[1-F^*(\varepsilon\lambda\sigma^{-1})], \qquad -\infty \leqq \varepsilon \leqq +\infty, \tag{8}$$

where $\sigma^2 = \sigma_u^2+\sigma_v^2$, $\lambda = \sigma_u/\sigma_v$, and $f^*(\cdot)$ and $F^*(\cdot)$ are the standard normal density and distribution functions, respectively. This density is asymmetric around zero, with its mean and variance given by

$$E(\varepsilon) = E(u) = -\frac{\sqrt{2}}{\sqrt{\Pi}}\sigma_u$$

$$\begin{aligned} V(\varepsilon) &= V(u)+V(v) \\ &= \left(\frac{\Pi-2}{\Pi}\right)\sigma_u^2+\sigma_v^2, \end{aligned} \tag{9}$$

as can be easily ascertained from elementary considerations and calculation of the moments of u.

The particular parameterization in (8) is convenient because λ is thereby interpreted to be an indicator of the relative variability of the two sources of random error that distinguish firms from one another.[4] $\lambda^2 \to 0$ implies $\sigma_v^2 \to \infty$ and/or $\sigma_u^2 \to 0$, i.e. that the symmetric error dominates in the determination of ε. Eq. (8) then becomes the density of a $N(0, \sigma^2)$ random variable, as can be seen by inspection. Similarly, when $\sigma_v^2 \to 0$, the one-sided error becomes the dominant source of random variation in the model and (5) takes on the form of a negative half-normal.[5]

[4]We prefer to use this interpretation of λ even though $\sigma_u{}^2$ is *not* the variance of $((\Pi-2)/\Pi)\,\sigma_u{}^2$ is. Another useful parameterization is to use σ^2 along with $\alpha = \sigma_u{}^2/\sigma^2$.

[5]For $\sigma_v{}^2 = 0$ and thus $\lambda = \infty$, (5) becomes

$$\begin{aligned} f(\varepsilon) &= \frac{\sqrt{2}}{\sqrt{\Pi}\,\sigma_u}\exp\left(-\frac{1}{2\sigma_u{}^2}\varepsilon^2\right), && \text{for } \varepsilon \leqq 0, \\ &= 0, && \text{otherwise.} \end{aligned}$$

The estimation problem is posed by assuming we have available a random sample of N observations and then forming the relevant log-likelihood function,

$$\ln \mathscr{L}(y|\boldsymbol{\beta}, \lambda, \sigma^2) = N \ln \frac{\sqrt{2}}{\sqrt{\Pi}} + N \ln \sigma^{-1} + \sum_{i=1}^{N} \ln [1 - F^*(\varepsilon_i \lambda \sigma^{-1})] - \frac{1}{2\sigma^2} \sum_{i=1}^{N} \varepsilon_i^2, \tag{10}$$

which is almost exactly the form of the likelihood function considered by Amemiya (1973, p. 1015, eq. (10.2)).

Taking derivatives,

$$\frac{\partial \ln \mathscr{L}}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{N} (y_i - \boldsymbol{\beta}'\boldsymbol{x}_i)^2 + \frac{\lambda}{2\sigma^3} \sum_{i=1}^{N} \frac{f_i^*}{(1-F_i^*)} (y_i - \boldsymbol{\beta}'\boldsymbol{x}_i), \tag{11}$$

$$\frac{\partial \ln \mathscr{L}}{\partial \lambda} = -\frac{1}{\sigma} \sum_{i=1}^{N} \frac{f_i^*}{(1-F_i^*)} (y_i - \boldsymbol{\beta}'\boldsymbol{x}_i), \tag{12}$$

$$\frac{\partial \ln \mathscr{L}}{\partial \boldsymbol{\beta}} = \frac{1}{\sigma^2} \sum_{i=1}^{N} (y_i - \boldsymbol{\beta}'\boldsymbol{x}_i)\boldsymbol{x}_i + \frac{\lambda}{\sigma} \sum_{i=1}^{N} \frac{f_i^*}{(1-F_i^*)} \boldsymbol{x}_i, \tag{13}$$

where $\boldsymbol{x}_i$ is a $(k \times 1)$ vector consisting of elements in the ith row of X, and f_i^* and F_i^* are, respectively, the standard normal density and distribution functions evaluated at $(y_i - \boldsymbol{\beta}'\boldsymbol{x}_i)\,\lambda\sigma^{-1}$.

Given (12), we have that

$$\sum_{i=1}^{N} \frac{f_i^*}{(1-F_i^*)} (y_i - \boldsymbol{\beta}'\boldsymbol{x}_i) = 0,$$

at the optimum. Inserting this result into (11), the ML estimator for σ^2 is determined through

$$-\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{N} (y_i - \boldsymbol{\beta}'\boldsymbol{x}_i) = 0, \tag{14}$$

which yields

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^{N} (y_i - \boldsymbol{\beta}'\boldsymbol{x}_i)^2, \tag{15}$$

the basis for the usual ML estimator of residual variance in a regression model. But the determination of $\hat{\boldsymbol{\beta}}$ is not independent of $\hat{\sigma}^2$ from other equations. In any event, this result can be used as a basis for an iterative solution scheme.

$\boldsymbol{\beta}'$ premultiplied into (13) gives

$$\frac{1}{\sigma^2}\sum_{i=1}^{N}(y_i-\boldsymbol{\beta}'\boldsymbol{x}_i)\,\boldsymbol{\beta}'\boldsymbol{x}_i+\frac{\lambda}{\sigma}\sum_{i=1}^{N}\frac{f_i^*}{(1-F_i^*)}\,\boldsymbol{\beta}'\boldsymbol{x}_i=0. \tag{16}$$

Adding to this $-\lambda$ times eq. (12) and simplifying, we get

$$\frac{1}{\sigma^2}\sum_{i=1}^{N}(y_i-\boldsymbol{\beta}'\boldsymbol{x}_i)\,\boldsymbol{\beta}'\boldsymbol{x}_i+\frac{\lambda}{\sigma}\sum_{i=1}^{N}\frac{f_i^*}{(1-F_i^*)}\,y_i=0, \tag{17}$$

which, in conjunction with (13) gives a system of $(k+1)$ equations that corresponds very closely to the system of first-order equations encountered in the so-called 'Tobit' model.[6] Since our density function is continuous in the range of ε, it is not anticipated that the difficulties encountered in the Tobit model will occur here, and therefore we claim all the usual maximum-likelihood properties for the values of $\boldsymbol{\beta}$, λ and σ^2 which simultaneously equate (11), (12), and (13) to zero. Formal proof of this claim and an examination of the regularity conditions that support it follow the analysis in Amemiya (1973).

Various solution algorithms are available for finding the optimizing values of $\boldsymbol{\beta}$, λ, and σ^2. Most of these (the Fletcher–Powell algorithm, for example) require analytical first- or second-order derivatives in addition to the likelihood function itself for their best performance at reasonable cost in computer time. Since such algorithms are now readily available,[7] we will not devote any space to a discussion of the ML computational problem, except to note that this likelihood function seems to be well-behaved, based on our experience. Second-order derivatives are presented in the appendix to this paper, for that use and as a basis for calculating asymptotic standard errors of the ML estimates.

We note in passing that if estimation of $\boldsymbol{\beta}$ alone is desired, all but the coefficient in $\boldsymbol{\beta}$ corresponding to a column of ones in X is estimated unbiasedly and consistently by least squares. Moreover, the components of σ^2 can be extracted (i.e., consistent estimators for them can be found) based on the least squares results by utilizing eq. (9) for $V(\varepsilon)$ in terms of σ_u^2 and σ_v^2 and a similar relationship for a higher-order moment of ε, since $V(\varepsilon)$ and higher order mean-corrected

[6]See Amemiya (1973, p. 1011, eqs. (7.2), (7.3)).

[7]For a good discussion of the available algorithms as of a few years ago, see Goldfeld–Quandt (1971).

moments of ε are themselves consistently estimable from the computed least-squares residuals.[8]

Similar comments and derivations would apply under alternative distributional assumptions for u_i. For example, we could choose the simple one-parameter exponential distribution for $-u$,

$$f(u) = \frac{1}{\phi}\exp(u/\phi), \qquad u \leqq 0, \tag{18}$$

where $\phi \geqq 0$ is the mean of $-u_i$. (The variance is ϕ^2.) A little algebra reveals that the distribution of $\varepsilon_i = v_i + u_i$ is given by the density

$$f(\varepsilon) = \frac{1}{\phi}\left[1 - F^*\left(\frac{\varepsilon}{\sigma_v} + \frac{\sigma_v}{\phi}\right)\right]\exp\left[\frac{\varepsilon}{\phi} + \frac{\sigma_v^2}{2\phi^2}\right], \tag{19}$$

where again $F^*(\cdot)$ represents the cumulative distribution function of the standard normal distribution. The likelihood function for the model follows immediately.

5. Monte Carlo results

In order to discover some specific information about the small sample behavior of the ML estimators discussed in the previous section, we constructed two limited Monte Carlo experiments which rest entirely on artificial data.

Study 1. The model considered here is $y_i = \varepsilon_i$ $(i = 1, \ldots, N)$ where ε_i is generated by eq. (8) with various values of σ_u^2 and σ_v^2. The results and further details about the range of parameter values considered are reported in table 1. So, no regression is involved in this instance (indeed, not even a mean is estimated) and we assume that the conventionally unobservable ε_i is observable.

Some easily distinguished patterns of bias and precision in estimation emerge. Considering table 1a, $\hat{\lambda}$ is apparently biased upward. σ_v^2, σ_u^2, and σ^2 seems to be estimated very well in all cases, both with regard to small bias and MSE.

Moving to table 1b, increasing sample size to $N = 100$ for the one set of results reported shows little (if any) basis for alteration of the qualitative conclusions reached above, but also gives to indication of the reduction in bias and increase in precision promised by the asymptotic properties of ML.

[8]For instance, the third-order moment of ε is

$$E[\varepsilon - E(\varepsilon)]^3 = \frac{2\sigma_u{}^3}{\sqrt{2\Pi}}\left(1 - \frac{4}{\Pi}\right).$$

In the Monte Carlo results that follow, no attempt is made to evaluate the properties of these estimators.

Table 1a

Monte Carlo results for the model $y_i = \varepsilon_i$; number of replications = 100, sample size = 50, value of λ = 1.66, 1.24, 0.83.[a]

1.66	1.24	0.83
(1.41) $\hat{\sigma}^2$ = 1.36 (0.08)	(1.22) $\hat{\sigma}^2$ = 1.22 (0.06)	(1.01) $\hat{\sigma}^2$ = 1.02 (0.04)
$\hat{\lambda}$ = 1.77 (0.28)	$\hat{\lambda}$ = 1.31 (0.20)	$\hat{\lambda}$ = 0.93 (0.13)
(1.03) $\hat{\sigma}_u{}^2$ = 1.01 (0.07)	(0.74) $\hat{\sigma}_u{}^2$ = 0.75 (0.07)	(0.41) $\hat{\sigma}_u{}^2$ = 0.47 (0.05)
(0.38) $\hat{\sigma}_v{}^2$ = 0.36 (0.02)	(0.48) $\hat{\sigma}_v{}^2$ = 0.47 (0.02)	(0.60) $\hat{\sigma}_v{}^2$ = 0.55 (0.03)
(1.88) $\hat{\sigma}^2$ = 1.81 (0.14)	(1.63) $\hat{\sigma}^2$ = 1.64 (0.11)	(1.35) $\hat{\sigma}^2$ = 1.36 (0.08)
$\hat{\lambda}$ = 1.80 (0.33)	$\hat{\lambda}$ = 1.35 (0.17)	$\hat{\lambda}$ = 0.94 (0.13)
(1.38) $\hat{\sigma}_u{}^2$ = 1.34 (0.16)	(0.99) $\hat{\sigma}_u{}^2$ = 1.03 (0.11)	(0.55) $\hat{\sigma}_u{}^2$ = 0.62 (0.09)
(0.50) $\hat{\sigma}_v{}^2$ = 0.47 (0.03)	(0.64) $\hat{\sigma}_v{}^2$ = 0.61 (0.04)	(0.80) $\hat{\sigma}_v{}^2$ = 0.74 (0.04)
(2.81) $\hat{\sigma}^2$ = 2.81 (0.33)	(2.45) $\hat{\sigma}^2$ = 2.46 (0.30)	(2.03) $\hat{\sigma}^2$ = 2.05 (0.19)
$\hat{\lambda}$ = 1.76 (0.35)	$\hat{\lambda}$ = 1.34 (0.22)	$\hat{\lambda}$ = 0.84 (0.08)
(2.06) $\hat{\sigma}_u{}^2$ = 2.06 (0.42)	(1.49) $\hat{\sigma}_u{}^2$ = 1.53 (0.30)	(0.83) $\hat{\sigma}_u{}^2$ = 0.85 (0.18)
(0.75) $\hat{\sigma}_v{}^2$ = 0.75 (0.07)	(0.96) $\hat{\sigma}_v{}^2$ = 0.93 (0.10)	(1.20) $\hat{\sigma}_v{}^2$ = 1.20 (0.09)

[a]Values in parentheses to the right of estimates are MSE's. Values in parentheses to the left of $\hat{\sigma}^2$, $\hat{\sigma}_u{}^2$ and $\hat{\sigma}_v{}^2$ are the true values for these parameters used in the simulations.

Table 1b

Monte Carlo results for the model $y_i = \varepsilon_i$; number of replications = 100, sample size = 100, value of λ = 1.66, 1.24, 0.83.[a]

1.66	1.24	0.83
(1.88) $\hat{\sigma}^2$ = 1.88 (0.14)	(1.63) $\hat{\sigma}^2$ = 1.70 (0.12)	(1.35) $\hat{\sigma}^2$ = 1.32 (0.08)
$\hat{\lambda}$ = 1.81 (0.37)	$\hat{\lambda}$ = 1.34 (0.23)	$\hat{\lambda}$ = 0.90 (0.12)
(1.38) $\hat{\sigma}_u{}^2$ = 1.40 (0.15)	(0.99) $\hat{\sigma}_u{}^2$ = 1.06 (0.14)	(0.55) $\hat{\sigma}_u{}^2$ = 0.57 (0.07)
(0.50) $\hat{\sigma}_v{}^2$ = 0.48 (0.03)	(0.64) $\hat{\sigma}_v{}^2$ = 0.64 (0.04)	(0.80) $\hat{\sigma}_v{}^2$ = 0.75 (0.05)

[a]Values in parentheses to the right of estimates are MSE's. Values in parentheses to the left of $\hat{\sigma}^2$, $\hat{\sigma}_u{}^2$ and $\hat{\sigma}_v{}^2$ are the true values for these parameters used in the simulations.

Study 2. Here the experiments of table 1 are repeated using the model $y_i = \mu + \varepsilon_i$ $(i = 1, \ldots, N)$. Since ε_i already possesses a non-zero mean, the question is, what effect does the extraction of an explicit intercept term have on the previous case? From table 2a we see that qualitatively the results for $\hat{\lambda}$ are altered somewhat, with some cases of negative bias occurring. Moreover, the additional parameter to be estimated has a deleterious effect on $\hat{\sigma}_u^2$ and (hence) $\hat{\sigma}^2$. Whereas in the previous cases these estimators provided sharp results, now there is a tendency for them both to be biased downward. $\hat{\mu}$ is also biased downward, but without any apparent pattern of relationship with the values of other parameters. Again, increasing sample size to $N = 100$ has no perceptible effect on our results and conclusions from the $N = 50$ case.

Table 2a

Monte Carlo results for the model $y_t = \mu + \varepsilon_t$; number of replications = 100, sample size = 50 μ = 1.0, value of λ = 1.66, 1.24, 0.83.[a]

1.66	1.24	0.83
(1.41) $\hat{\sigma}^2$ = 1.34 (0.22)	(1.22) $\hat{\sigma}^2$ = 1.11 (0.25)	(1.01) $\hat{\sigma}^2$ = 0.82 (0.13)
$\hat{\lambda}$ = 1.72 (1.22)	$\hat{\lambda}$ = 1.04 (1.20)	$\hat{\lambda}$ = 0.23 (0.66)
$\hat{\mu}$ = 0.88 (0.11)	$\hat{\mu}$ = 0.75 (0.21)	$\hat{\mu}$ = 0.55 (0.25)
(1.03) $\hat{\sigma}_u{}^2$ = 0.95 (0.34)	(0.74) $\hat{\sigma}_u{}^2$ = 0.58 (0.44)	(0.41) $\hat{\sigma}_u{}^2$ = 0.12 (0.23)
(0.38) $\hat{\sigma}_v{}^2$ = 0.39 (0.04)	(0.48) $\hat{\sigma}_v{}^2$ = 0.53 (0.05)	(0.60) $\hat{\sigma}^2$ = 0.70 (0.04)
(1.88) $\hat{\sigma}^2$ = 1.75 (0.39)	(1.63) $\hat{\sigma}^2$ = 1.61 (0.51)	(1.35) $\hat{\sigma}^2$ = 1.30 (0.30)
$\hat{\lambda}$ = 1.93 (2.99)	$\hat{\lambda}$ = 1.45 (2.73)	$\hat{\lambda}$ = 0.70 (0.91)
$\hat{\mu}$ = 0.84 (0.18)	$\hat{\mu}$ = 0.82 (0.25)	$\hat{\mu}$ = 0.71 (0.23)
(1.38) $\hat{\sigma}_u{}^2$ = 1.25 (0.65)	(0.99) $\hat{\sigma}_u{}^2$ = 1.00 (0.83)	(0.55) $\hat{\sigma}_u{}^2$ = 0.50 (0.52)
(0.50) $\hat{\sigma}_v{}^2$ = 0.50 (0.07)	(0.64) $\hat{\sigma}_v{}^2$ = 0.61 (0.09)	(0.80) $\hat{\sigma}_v{}^2$ = 0.80 (0.10)
(.81) $\hat{\sigma}^2$ = 2.56 (0.89)	(2.45) $\hat{\sigma}^2$ = 2.40 (0.93)	(2.03) $\hat{\sigma}^2$ = 2.02 (0.48)
$\hat{\lambda}$ = 1.84 (2.39)	$\hat{\lambda}$ = 1.46 (1.32)	$\hat{\lambda}$ = 0.81 (0.79)
$\hat{\mu}$ = 0.83 (0.23)	$\hat{\mu}$ = 0.85 (0.29)	$\hat{\mu}$ = 0.78 (0.25)
(2.06) $\hat{\sigma}_u{}^2$ = 1.80 (1.35)	(1.49) $\hat{\sigma}_u{}^2$ = 1.55 (1.48)	(0.83) $\hat{\sigma}_u{}^2$ = 0.86 (1.05)
(0.75) $\hat{\sigma}_v{}^2$ = 0.76 (0.14)	(0.96) $\hat{\sigma}_v{}^2$ = 0.86 (0.18)	(1.20) $\hat{\sigma}_v{}^2$ = 1.16 (0.22)

[a]Values in parentheses to the right of estimates are MSE's. Values in parentheses to the left of $\hat{\sigma}^2$, $\hat{\sigma}_u{}^2$ and $\hat{\sigma}_v{}^2$ are the true values for these parameters used in the simulations.

Table 2b

Monte Carlo results for the model $y_t = \mu + \varepsilon_t$; number of replications = 100, sample size = 100, μ = 1.0, value of λ = 1.66, 1.24, 0.83.[a]

1.66	1.24	0.83
(1.88) $\hat{\sigma}^2$ = 1.90 (0.47)	(1.63) $\hat{\sigma}^2$ = 1.79 (0.55)	(1.35) $\hat{\sigma}^2$ = 1.17 (0.22)
$\hat{\lambda}$ = 2.03 (2.48)	$\hat{\lambda}$ = 1.57 (1.55)	$\hat{\lambda}$ = 0.43 (0.61)
$\hat{\mu}$ = 0.88 (0.17)	$\hat{\mu}$ = 0.93 (0.19)	$\hat{\mu}$ = 0.62 (0.24)
(1.38) $\hat{\sigma}_u{}^2$ = 1.40 (0.79)	(0.99) $\hat{\sigma}_u{}^2$ = 1.22 (0.87)	(0.55) $\hat{\sigma}_u{}^2$ = 0.29 (0.39)
(0.50) $\hat{\sigma}_v{}^2$ = 0.50 (0.08)	(0.64) $\hat{\sigma}_v{}^2$ = 0.57 (0.08)	(0.80) $\hat{\sigma}_v{}^2$ = 0.87 (0.07)

[a]Values in parentheses to the right of estimates are MSE's. Values in parentheses to the left of $\hat{\sigma}^2$, $\hat{\sigma}_u{}^2$ and $\hat{\sigma}_v{}^2$ are the true values for these parameters used in the simulations.

Contrasting these results to similar experiments conducted by the Aigner, Amemiya, and Poirier (1976), the overall performance of common estimators ($\hat{\mu}$ and $\hat{\sigma}^2$) is roughly comparable.[9] We note that their measure of the relative variability in error sources is generally overstated, but has small MSE compared to $\hat{\lambda}$.

[9]Cf. their table 5 on page 20. The disturbance variance, $V(\varepsilon)$, is set at 0.5 in their results, whereas it varies in our table 2a and is generally much larger than 0.5.

6. Empirical examples

In this section we present some examples of estimates of the stochastic frontier model, and compare these estimates to other earlier results.

The first example uses 1957–58 data on the U.S. primary metals industry (SIC 33), consisting of observations across 28 states. This data set was previously analyzed by Hildebrand and Liu (1965) and by Aigner and Chu (1968). The production function to be estimated is of the form

$$\ln Y = \beta_0+\beta_1 \ln L+\beta_2 (\ln R\cdot \ln K)+(v+u), \tag{20}$$

where Y is value added per establishment, L is a measure of labor input per establishment, K is the gross book value of plant and equipment per establishment, and R is the ratio of net to gross book value of plant and equipment.

Various sets of parameter estimates are given in table 3. The first three sets of entries correspond to our results running OLS, and using the maximum-likelihood technique of section 4 (with the one-sided disturbance assumed to be half-normal and exponential, respectively). The numbers in parentheses under the OLS estimates are t ratios. The numbers in parentheses under the stochastic frontier maximum-likelihood estimates are 'asymptotic t ratios'. That is, they are the ratio of the coefficient estimate to the square root of the appropriate diagonal element of the inverse of the information matrix. These are asymptotically distributed as $N(0, 1)$ under the null hypothesis that the associated coefficient is zero.

Table 3
Estimates of eq. (20) by various methods.

Method	β_0	β_1	β_2
OLS	0.9146 (2.04)	0.9168 (7.31)	0.04164 (2.19)
Stochastic frontier (exponential)	0.9601 (2.20)	0.9144 (7.71)	0.04125 (2.29)
Stochastic frontier (half-normal)	0.9600 (2.06)	0.9105 (7.68)	0.04208 (2.34)
Hildebrand–Liu OLS		0.988	0.04208
Aigner–Chu OLS		0.908	0.0333
Aigner–Chu LP		0.873	0.0031
Aigner–Chu QP1		1.071	0.0269
Aigner–Chu QP2		0.822	0.0219

The remaining entries are taken from Aigner and Chu (1968, p. 836). They consist of the OLS results of Hildebrand and Liu and of Aigner and Chu, the linear programming results of Aigner and Chu, and the results of two variants of the quadratic programming technique of Aigner and Chu. We did not succeed in duplicating exactly the OLS results of either Hildebrand and Liu or Aigner and Chu, although the discrepancies are not large.

What is most interesting is the extremely close agreement between our OLS estimates and both sets of stochastic frontier maximum-likelihood estimates. In particular, the stochastic frontier estimates are much closer to the OLS estimates than they are to Aigner and Chu's programming results.

The reason for this is clear if we look at the estimates of the parameters of the distributions of the disturbances.[10] In the OLS case the estimated variance of the disturbances is 0.077640. In the half-normal case we have

$$\hat{\sigma}_u^2 = 0.000686, \qquad \text{'asymptotic } t \text{ ratio'} = 0.05,$$
$$\hat{\sigma}_v^2 = 0.0692, \qquad \text{'asymptotic } t \text{ ratio'} = 3.64;$$

and in the exponential case we have

$$\hat{\phi} = 0.0180, \qquad \text{'asymptotic } t \text{ ratio'} = 0.16,$$
$$\hat{\sigma}_v^2 = 0.0691, \qquad \text{'asymptotic } t \text{ ratio'} = 3.68.$$

In both cases the symmetric component of the disturbance effectively swamps the one-sided component. In the exponential case the mean of the one-sided component is 0.018 and its variance is 0.000325, which is only 0.468 percent of the total disturbance variance of 0.000325+0.0691 = 0.0694. Similarly, in the half-normal case, the mean of the one-sided component is 0.021 and its variance is 0.000251, which is only 0.361 percent of the total disturbance variance. Therefore, the picture that emerges is one of substantial variation in the frontier across states, but relatively little variation of observed output beneath the frontier.

As a final note, the maximized value of the logarithm of the likelihood function is −2.372 in the exponential case, and −2.368 in the half-normal case, indicating a marginally better fit by the half-normal distribution.

Our second example uses U.S. agricultural data for six years (1960–65) and the 48 contiguous states. The data set is the one used by Timmer (1971), excluding the years 1966 and 1967. The function to be estimated is

$$Y = \beta_0+\beta_1 \text{ Labor}+\beta_2 \text{ Capital}+\beta_3 \text{ Land}+\beta_4 \text{ Fertilizer} \qquad (21)$$
$$+\beta_5 \text{ Livestock}+\beta_6 \text{ Seed}+(v+u),$$

[10]In the empirical examples used here, the likelihood function was explicitly parameterized in terms of the parameters of the two error component distributions, $\sigma_u{}^2$ and $\sigma_v{}^2$, rather than in the (equivalent) way discussed in section 4.

where Y is gross agricultural output and the remaining variables are (more or less) self-explanatory; see Timmer (1971) for details. All variables are in logarithms, and are on a per-farm basis.

Various sets of parameter estimates are given in table 4. The first two sets of estimates are our OLS and stochastic frontier results, based on an exponential distribution for $-u$. The other four sets are taken from Timmer (1971, p. 785). They represent, respectively, his OLS results, his linear programming results, and his linear programming results using 98 percent and 97 percent of the observations.

Table 4
Estimates of eq. (21) by various methods.

Method	β_0	β_1	β_2	β_3	β_4	β_5	β_6
OLS	1.8072	0.1149	0.2976	0.06061	0.1411	0.2581	0.1956
		(3.67)	(8.85)	(4.81)	(12.40)	(16.72)	(6.44)
Stochastic frontier	1.8143	0.1148	0.2776	0.06061	0.1411	0.2581	0.1956
(exponential)	(42.9)	(3.71)	(8.96)	(4.87)	(12.56)	(16.93)	(6.52)
Timmer OLS	1.7350	0.1919	0.3726	0.0458	0.1484	0.2510	0.1579
	(53.8)	(6.7)	(11.7)	(4.2)	(16.0)	(19.5)	(5.4)
Timmer LP_{100}	1.6693	0.6015	0.4887	-	0.1334	0.2347	0.1043
Timmer LP_{98}	1.8578	0.3287	0.3689	0.0298	0.1428	0.2045	0.2243
Timmer LP_{97}	1.8828	0.2679	0.4842	0.0099	0.1693	0.1885	0.1712

Our OLS results are probably as close as can be expected to Timmer's OLS results, given that we are missing 96 of his observations. Our stochastic frontier estimates approximate the OLS estimates, as in the previous example. They are also reasonably close to Timmer's OLS, LP_{98} and LP_{97} estimates. It would appear that the linear programming (LP_{100}) estimates are the only ones very unlike the others.

Our estimates of the parameters of the distributions of the disturbances are

$$\hat{\phi} = 0.00710, \qquad \text{'asymptotic } t \text{ ratio'} = 0.30,$$

$$\hat{\sigma}_v^2 = 0.01005, \qquad \text{'asymptotic } t \text{ ratio'} = 11.2$$

Once again the symmetric component of the disturbance completely swamps the exponential component; the variance of the exponential component (0.0000504) is only about one half of one percent of the total variance.

7. Conclusions

We have described a linear model with an error specification that is considered appropriate for the estimation of an industry production function using cross-section data. The specification does not prejudge the placement of the function, as if an 'average' function or a 'frontier' function were to be fitted to the data. Indeed, the interesting feature of the framework discussed herein is precisely that placement of the function is estimated along with other model parameters. This is also the contribution of the model presented in Aigner, Amemiya, and Poirier (1976), but that model is not capable of direct interpretation in terms of the sources of random error that may cause firms with identical input vectors to differ.

Whether, from an applications viewpoint, one model dominates the other is as yet not clear, since the empirical evidence from both studies is limited. What we find in the Monte Carlo results of the present paper is not particularly encouraging: In the presence of an intercept, the 'placement' parameter, λ, and the intercept itself are apparently difficult to estimate by the ML technique even when sample size is as large as 100. Further small-sample studies would help in forming a more definite opinion about the relative merits of the alternative specifications for the placement and estimation of the industry production studied by Aigner, Amemiya, and Poirier, and in the present paper. Additional research is also required to evaluate the performance of the 'moment' estimators based on least-squares residuals mentioned at the end of section 4.

Tests of our model on two real-world data sets indicated relatively small one-sided components of the disturbance. This in turn suggests high levels of efficiency relative to a stochastic frontier. Whether this finding, based on statewide per-establishment aggregates, would continue to hold for the individual establishments themselves is yet another interesting question to be answered.

Appendix

Below are reproduced the second-order derivatives of $\ln \mathscr{L}$ with respect to $\boldsymbol{\beta}$, λ, and σ^2 for use in computational algorithms and as a basis for estimating asymptotic standard errors,

$$\frac{\partial^2 \ln \mathscr{L}}{\partial \lambda^2} = \frac{1}{\sigma^2} \sum_{i=1}^{N} \frac{f_i^*}{(1-F_i^*)^2} (y_i - \boldsymbol{\beta}'\boldsymbol{x}_i)^2 \left\{ -f_i^* + \frac{\lambda}{\sigma}(1-F_i^*)(y_i - \boldsymbol{\beta}'\boldsymbol{x}_i) \right\} \tag{A.1}$$

$$\frac{\partial^2 \ln \mathscr{L}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = -\frac{1}{\sigma^2} \sum_{i=1}^{N} \boldsymbol{x}_i \boldsymbol{x}_i' + \frac{\lambda^2}{\sigma^2} \sum_{i=1}^{N} \frac{f_i^*}{(1-F_i^*)^2} \left\{ -f_i^* + \frac{\lambda}{\sigma}(1-F_i^*)(y_i - \boldsymbol{\beta}'\boldsymbol{x}_i) \right\} \boldsymbol{x}_i \boldsymbol{x}_i', \tag{A.2}$$

$$\frac{\partial^2 \ln \mathscr{L}}{\partial(\sigma^2)^2} = \frac{N}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{N}(y_i-\boldsymbol{\beta}'\boldsymbol{x}_i)^2 + \frac{\lambda}{4\sigma^5}\sum_{i=1}^{N}\frac{f_i^*}{(1-F_i^*)^2}\left\{-\frac{\lambda}{\sigma}f_i^*(y_i-\boldsymbol{\beta}'\boldsymbol{x}_i)^2 + \frac{\lambda^2}{\sigma^2}(1-F_i^*)(y_i-\boldsymbol{\beta}'\boldsymbol{x}_i)^3 - 3(1-F_i^*)(y_i-\boldsymbol{\beta}'\boldsymbol{x}_i)\right\}, \tag{A.3}$$

$$\frac{\partial^2 \ln \mathscr{L}}{\partial\lambda\partial\boldsymbol{\beta}} = -\frac{1}{\sigma}\sum_{i=1}^{N}\frac{f_i^*}{(1-F_i^*)^2}\left\{-(1-F_i^*) - \frac{\lambda}{\sigma}f_i^*(y_i-\boldsymbol{\beta}'\boldsymbol{x}_i) + \frac{\lambda^2}{\sigma^2}(1-F_i^*)(y_i-\boldsymbol{\beta}'\boldsymbol{x}_i)^2\right\}\boldsymbol{x}_i, \tag{A.4}$$

$$\frac{\partial^2 \ln \mathscr{L}}{\partial\sigma^2\partial\boldsymbol{\beta}} = -\frac{1}{\sigma^4}\sum_{i=1}^{N}(y_i-\boldsymbol{\beta}'\boldsymbol{x}_i)\boldsymbol{x}_i + \frac{\lambda}{2\sigma^3}\sum_{i=1}^{N}\frac{f_i^*}{(1-F_i^*)^2}\left\{-(1-F_i^*) - \frac{\lambda}{\sigma}f_i^*(y_i-\boldsymbol{\beta}'\boldsymbol{x}_i) + \frac{\lambda^2}{\sigma^2}(1-F_i^*)(y_i-\boldsymbol{\beta}'\boldsymbol{x}_i)^2\right\}\boldsymbol{x}_i, \tag{A.5}$$

$$\frac{\partial^2 \ln \mathscr{L}}{\partial\sigma^2\partial\lambda} = \frac{1}{2\sigma^3}\sum_{i=1}^{N}\frac{f_i^*}{(1-F_i^*)^2}\left\{(1-F_i^*)(y_i-\boldsymbol{\beta}'\boldsymbol{x}_i) + \frac{\lambda}{\sigma}f_i^*(y_i-\boldsymbol{\beta}'\boldsymbol{x}_i)^2 - \frac{\lambda^2}{\sigma^2}(1-F_i^*)(y_i-\boldsymbol{\beta}'\boldsymbol{x}_i)^3\right\}. \tag{A.6}$$

References

Afriat, S.N., 1972, Efficiency estimation of production functions, International Economic Review 13, Oct., 568–598.

Aigner, D.J. and S.F. Chu, 1968, On estimating the industry production function, American Economic Review 58, 826–839.

Aigner, D.J., T. Amemiya and D.J. Poirier, 1976, On the estimation of production frontiers, International Economic Review 17, 377–396.

Amemiya, Takeshi, 1973, Regression analysis when the dependent variable is truncated normal, Econometrica 41, 997–1016.

Cohen, A.C., Jr., 1950, Estimating the mean and variance of normal populations from singly and doubly truncated samples, Annals of Mathematical Statistics 21, 557–569.

Cohen, A.C., Jr., 1957, On the solution of estimating equations for truncated and censored samples from normal populations, Biometrika 44, 225–236.

Dugger, R., 1974, An application of bounded nonparametric estimating functions to the analysis of bank cost and production functions, unpublished Ph.D. dissertation (University of North Carolina, Chapel Hill, NC).

Farrell, M.J., 1957, The measurement of productive efficiency, Journal of the Royal Statistical Society (A, general) 120, pt. 3, 253–281.

Goldfeld, Stephen and Richard Quandt, 1971, Nonlinear methods in econometrics (North-Holland, Amsterdam).

Halperin, M., 1952, Maximum likelihood estimation in truncated samples, Annals of Mathematical Statistics 23, 226–238.

Halperin, M., 1952, Estimation in the truncated normal distribution, Journal of the American Statistical Association 47, 457–465.

Hartley, H.O., 1958, Maximum likelihood estimation from incomplete data, Biometrics 14, 174–194.

Hildebrand, G. and T.C. Liu, 1965, Manufacturing production functions in the United States, 1957 (Cornell University Press, Ithaca, NY).

Marschak, J. and W.J. Andrews, 1944, Random simultaneous equations and the theory of production, Econometrica 12, July–Oct., 143–205.

Richmond, J., 1974, Estimating the efficiency of production, International Economic Review 15, June, 515–521.

Schmidt, P., 1976, On the statistical estimation of parametric frontier production functions, Review of Economics and Statistics 58, 238–239.

Timmer, C. Peter, 1971, Using a probabilistic frontier production function to measure technical efficiency, Journal of Political Economy 79, 776–794.

Weinstein, M.A., 1964, The sum of values from a normal and a truncated normal distribution, Technometrics 6, 104–105 (with some additional material, 469–470).

Zellner, A., Kmenta, J. and J. Drèze, 1966, Specification and estimation of Cobb–Douglas production functions, Econometrica 34, Oct., 784–795.