

Working Paper 92-11
March 1992

División de Economía
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)

SWITCHING REGRESSIONS AND ACTIVITY ANALYSIS

Eduardo Ley*

Abstract

We study the use of switching regression models to characterize the coefficients in linear production technologies with a finite number of activities. Maximum likelihood-based methods are proposed and different switching specifications are discussed. The viability of these newly proposed techniques is established. The methods developed combine the advantages of the two major approaches to frontier estimation: the functional flexibility of the linear programming—nonparametric and nonstatistical—approach and the statistical nature of the econometric—both parametric and statistical—approach. This combination comes at the expense of some analytical complexity.

Key Words

Switching regression models, activity analysis, linear production models.

* Departamento de Economía, Universidad Carlos III de Madrid.

Switching Regressions and Activity Analysis

Eduardo LPY

1. Introduction

In this paper we investigate the use of switching regression models in the estimation of production coefficients of linear activity technologies. Technologies with smoother isoquants can be thought of as the limit of a linear technology when the number of activities grows to infinity. We explore here an approach to production frontier estimation based on the direct study of these primary activities. The main advantage of this approach is its functional flexibility. Alternative econometric methods to frontier analysis postulate specific and quite restrictive parametric forms which often encounter problems fitting the observed data.

1.1. Frontier Estimation

There are two main approaches to frontier estimation: the *parametric* approach and the *nonparametric* approach (here, parametric and nonparametric refer to the form of the production or cost functions and not to the error specification). The parametric approach assumes particular parametric form for, say, the production function (*e.g.*, Cobb-Douglas, CES, translog), adds a one-sided¹ random error reflecting the presence of technical inefficiency in the production process, and, sometimes, a second stochastic component reflecting the effects of noise, measurement error, model misspecification and exogenous shocks. The nonparametric approach doesn't assume any specific form for the production function but *envelopes*² the sample data by the smallest convex weak-disposal hull that satisfies the axioms imposed by economic theory.³ Since very little structure is imposed on the data this approach is unable to accommodate stochastic elements in a satisfactory way. See Førsund, Lovell and Schmidt (1980) and Schmidt (1985) for a survey of all different methods of frontier estimation; see Bauer (1990) and Seiford and Thrall (1990) for surveys of recent developments of the parametric and nonparametric approaches to frontier estimation.

The main drawback of the parametric approach is that the particular parametric form used by the analyst is a maintained hypothesis imposed on the data which cannot be tested—*i.e.*, it's an act of faith. The nonparametric approach, on the other hand, is a data-based model building method; however its inability to accommodate random noise makes its results very sensitive to measurement error and model misspecification. There have been various attempts to combine the functional flexibility of the nonparametric approach with the ability of handling statistical noise.

I wish to thank Hal Varian and Margarida Genius who provided many useful discussions. Financial support from the Bank of Spain is gratefully acknowledged.

¹ Non-positive in the case of production frontiers, and non-negative in the case of cost frontiers.

² Hence the term *data envelopment analysis* is often used to refer to these techniques.

³ See Shephard (1970) for a list of the regularity conditions desired for production functions.

Varian (1985) incorporates measurement error to the nonparametric tests of optimizing behavior. However, the paper is concerned with testing, not with estimation, and no new methods for recovering technological parameters are developed.

Banker and Maindiratta (1987) and Banker (1988) have attempted to lay some statistical foundations for the nonparametric techniques known as *data envelopment analysis* (DEA). Assuming a monotonically decreasing density function for the deviation between actual and efficient level of output, Banker (1988) proves that the DEA estimates of the best-practice monotone-increasing and concave production function are also maximum likelihood estimates. Since the number of *incidental* parameters⁴ to be estimated by DEA methods grows with the sample size, the usual statistical properties of the maximum likelihood estimators do not apply there. However, Banker (1988) manages to prove the consistency of the DEA estimates from first principles. Nevertheless, other properties of these estimators are uncertain and no estimates for their standard errors are derived. Moreover, these papers are efforts to motivate DEA instead than attempts to develop an appropriate statistical framework for the nonparametric approach.

Finally, Land, Lovell and Thore (1988) use chance-constrained programming techniques to allow for uncertainty about the structure of the efficient production technology. They append the methods of chance-constrained programming to the nonparametric deterministic frontier model. In addition to the usual input-output data, evaluator-supplied information concerning accuracy of the data and willingness to take risk are required. Furthermore, chance-constrained efficiency measurement continues to be deterministic: efficiency is calculated by means of nonlinear programming techniques and no parameters are actually estimated in the process.

The approach that we present in this paper is based on a stochastic specification in a linear activities context which allows the use of statistical techniques to estimate technological parameters. Minimal functional constraints are imposed, as in the DEA approach. At the same time, a composed-error specification is borrowed from the econometric approach to frontier estimation.

2. Activity Analysis

2.1. Linear Technologies

Following Koopmans (1951), we characterize a firm's production technical possibilities by two basic concepts, commodities and activities. Each commodity is assumed to be of homogeneous quality and infinitely divisible. Commodities include primary factors of production, intermediate products and final products. An *activity* will consist of the combination of a number of commodities—inputs—in fixed quantities to produce other commodities—outputs. Each of the I different activities employs J inputs to produce the single output y (for simplicity, we won't consider multioutput technologies). The i th activity will be characterized by a set of coefficients,

$$\Gamma'_i \equiv (\gamma_{i1}, \gamma_{i2}, \dots, \gamma_{iJ}) \tag{1}$$

indicating the amount of each input needed to produce one unit of output (again, for the sake of simplicity, we won't consider joint production). We'll assume that the firm has

⁴ See Chamberlain (1980) for a discussion of the "incidental parameters problem."

access to a finite number of activities, I ; write $\Gamma = (\Gamma_1, \Gamma_2, \dots, \Gamma_I)$. Let x_{ij} be the amount of input j used in activity i , then $x_j = \sum_i x_{ij}$ is the total quantity of input j used in the production process; write $\mathbf{x}' = (x_1, x_2, \dots, x_J)$. Let w_j the unit price of input j , and put $\mathbf{w}' = (w_1, w_2, \dots, w_J)$.

Two assumptions are usually made on this concept of production activity: divisibility and additivity. Divisibility implies that each activity is capable of a continuous reduction or expansion. In particular it is assumed that $\mathbf{x} = k\Gamma_i$ can be used to produce k units of output, for any $k > 0$. It follows that divisibility implies constant returns to scale. Additivity means that any number of activities might be carried simultaneously without modification of the technical ratios defined by (1). Let θ_i be level of operation of activity i —i.e., θ_i units of output are obtained from activity i —which is given by

$$\theta_i = \min_j \left\{ \frac{x_{ij}}{\gamma_{ij}} \right\};$$

write $\Theta' = (\theta_1, \theta_2, \dots, \theta_I)$. Total output will be given by

$$y = \sum_{i=1}^I \theta_i.$$

The cost minimizing problem can be stated as

$$\begin{aligned} \min_{\Theta} \quad & \mathbf{w}'\Gamma\Theta \\ \text{s.t.} \quad & \mathbf{1}'\Theta \geq y \\ & \theta_i \geq 0 \quad \forall i \end{aligned}$$

where $\mathbf{1}_I$ is a column vector of I ones. Koopmans (1951) studied the existence and characterization of the solutions to this problem. The optimal solution—which doesn't need to be unique—can be more easily obtained by looking at the dual program

$$\begin{aligned} \max_{\phi} \quad & y\phi \\ \text{s.t.} \quad & \phi \leq \mathbf{w}'\Gamma_i \quad \forall i \\ & \phi \geq 0 \end{aligned}$$

which does always have a unique solution. It is easy to see that we'll have

$$y\phi^* = y \min_i \{\mathbf{w}'\Gamma_i\},$$

which is the minimum cost of producing y . More than one optimal intensity vector, Θ^* , might be associated with this minimum cost.

This paper addresses the problem of estimating the coefficients in Γ once data about output, y , and cost, $\mathbf{w}'\mathbf{x}$, are known. We also investigate the problem of determining the dimension of the matrix Γ —i.e., the issue of determining the number of activities present. As relative prices change, some activities become more economically attractive than others. We propose to use the prices in the vector \mathbf{w} as the regressors in a switching regression model. Firms will be *switching* from some activities to others in response to variations in \mathbf{w} . An endogenously switching regression model with sample separation unknown (to be introduced below) seems to offer an appropriate and promising approach to recovering the characteristics of the underlying technology.

2.2. Nonparametric Approach to Frontier Estimation as Activity Analysis

The axiomatic nonparametric approach to frontier estimation uses a sequence of linear programs to construct a transformation frontier and to compute efficiency measures relative to this reference technology. See Charnes, Cooper and Rhodes (1978), and Banker, Charnes and Cooper (1984) for a detailed description of these methods. This work is a generalization of the techniques developed in Farrell (1957); see Färe, Grosskopf and Lovell (1985) for a detailed exposition of different generalizations and alternative measures of the Farrell input-based measure of efficiency.

Suppose we have a set of observations about output, $\mathbf{y} = (y_1, y_2, \dots, y_n)$, and on J inputs $\mathbf{x}_n = (x_{1n}, x_{2n}, \dots, x_{Jn})$, where the index $n = 1, 2, \dots, N$, might refer to different decision units, or to different observations on the same firms. The objective is to construct a reference technology, and, then, to compare each unit n 's performance relative to the rest of the observations. The nonparametric methods solve, for each $m = 1, 2, \dots, N$, the following linear programs⁵

$$\begin{aligned} \max_{\psi_n} \quad & \sum_{n=1}^N \psi_n y_n \\ \text{s.t.} \quad & \sum_{n=1}^N \psi_n x_{jn} \leq x_{jm} \quad \forall j \\ & \psi_n \geq 0 \quad \forall n \end{aligned}$$

this problem is equivalent to

$$\begin{aligned} \max_{\psi_n} \quad & \sum_{n=1}^N \hat{\psi}_n \\ \text{s.t.} \quad & \sum_{n=1}^N \hat{\psi}_n \hat{\gamma}_{jn} \leq x_{jm} \quad \forall j \\ & \hat{\psi}_n \geq 0 \quad \forall n \end{aligned}$$

where $\hat{\psi}_n = \psi_n y_n$ and $\hat{\gamma}_{jn} = x_{jn}/y_n$. The interpretation is now straightforward; $\hat{\psi}_n$ is the intensity or level of operation of 'activity' n . Total output of unit m is being maximized subject to the constraint of not using up any more of each input j than what was actually observed to be utilized by unit m . 'Activity' n is characterized by the input-output list associated with unit n . Thus, for each unit m , we check whether, combining the techniques observed in other units, more output could have been obtained without using more inputs. If the answer is yes, then such unit is termed *inefficient*. The extent of the inefficiency can be measured by the ratio of the length of the input vector to the length of the maximum radial contraction of the input vector which can still produce the same level of output. Let $E \equiv \{n \in \{1, 2, \dots, N\} : \text{unit } n \text{ is efficient}\}$; then the recovered technology is characterized by $\hat{\Gamma}'_n \equiv (\hat{\gamma}_{1n}, \hat{\gamma}_{2n}, \dots, \hat{\gamma}_{Jn})$ such that $n \in E$.

⁵ For the sake of simplicity, we're assuming constant returns to scale here but more general specifications are possible with just a slight modification of the linear program—i.e., an additional constraint must be incorporated—; see, for example, Färe, Grosskopf and Lovell (1985).

In summary, these techniques determine the dimension of the matrix Γ by the number of efficient units, and its elements are calculated by the ratios $\hat{\gamma}_{jn} = x_{jn}/y_n$.⁶ The approach that we propose is based on a stochastic specification that allows the use of maximum likelihood-based methods to determine the dimension of the technology matrix, Γ , and to estimate its elements. Figure 1 illustrates the difference between these two approaches to recovering the underlying technology.

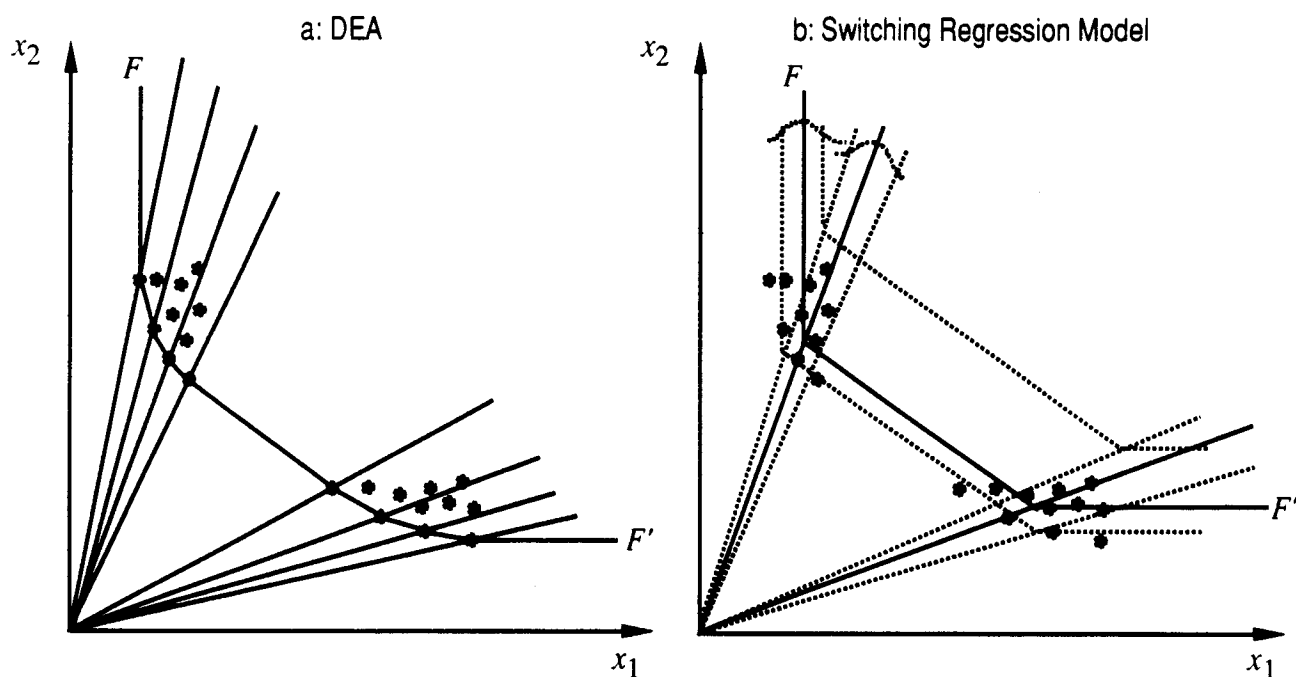


Figure 1. Two approaches to recovering the underlying technology.

In figure 1, we have plotted *hypothetically observed* input combinations all producing one unit of output. In 1.a, the nonparametric technique, DEA, involves the observations by the smallest convex hull. The ‘estimated’ isoquant, FF' , is piecewise linear with kinks at the ‘estimated’ activities’ expansions paths. The switching regression activity model, in 1.b, estimates a stochastic isoquant, FF' , with the associated estimated activity coefficients. DEA recovers eight different activities while the more parsimonious switching regression model estimates coefficients for only two activities. The broken lines in 1.b represent “confidence intervals.”

3. Switching Regression Models

We’ll give here a short review of switching regression models; Maddala (1986) can be consulted for a more detailed exposition. Another good early reference is Goldfeld and Quandt (1973).⁷ We start assuming that the observations of a dependent variable, y , may follow two different data-generating mechanisms in two different *regimes*, regime 1 and

⁶ It must be noted that no data on prices are needed in order to apply this method.

⁷ However, as noted in Maddala and Nelson (1974), the likelihood functions in the models in Goldfeld and Quandt (1973) didn’t take into account all the relevant information pertinent to the switch from one regime to another and are, therefore, misspecified

regime 2:

$$\begin{aligned} y_{1t} &= \mathbf{x}'_t \beta_1 + \varepsilon_{1t} & (\text{regime 1}) \\ y_{2t} &= \mathbf{x}'_t \beta_2 + \varepsilon_{2t} & (\text{regime 2}) \end{aligned} \quad (2)$$

and

$$y_t = \begin{cases} y_{1t} & \text{iff } \mathbf{z}'_t \alpha - \nu_t < 0 \\ y_{2t} & \text{iff } \mathbf{z}'_t \alpha - \nu_t > 0 \end{cases} \quad (3)$$

where \mathbf{x}'_t and \mathbf{z}'_t , for $t = 1, 2, \dots, T$, are non-exclusive sets of data, β_1 , β_2 and α are parameters to be estimated, and ε_{1t} , ε_{2t} and ν_t are residuals that are at most (possibly) contemporaneously correlated. Depending on the assumptions made on these correlations we might have *endogenous* switching (when $\text{Cov}[\varepsilon_{it}, \nu_t] \neq 0$) or *exogenous* switching (when $\text{Cov}[\varepsilon_{it}, \nu_s] = 0$, for all t and s). We have a model with *sample separation known* when we know which observations belong to either regime; otherwise it is a model with *sample separation unknown*. Equations (2) and (3) constitute a switching regression model.

The models that will arise below in the activity analysis context are endogenous switching regression models with sample separation unknown. In particular, we'll be looking at models of the form

$$\begin{aligned} y_{1t} &= \mathbf{x}'_t \beta_1 + \varepsilon_{1t} \\ y_{2t} &= \mathbf{x}'_t \beta_2 + \varepsilon_{2t} \end{aligned}$$

and

$$y_t = \begin{cases} y_{1t} & \text{iff } \mathbf{x}'_t (\beta_1 - \beta_2) + (\varepsilon_{1t} - \varepsilon_{2t}) < 0 \\ y_{2t} & \text{iff } \mathbf{x}'_t (\beta_1 - \beta_2) + (\varepsilon_{1t} - \varepsilon_{2t}) > 0 \end{cases}$$

which implies that $\mathbf{z}_t = \mathbf{x}_t$, $\alpha = \beta_1 - \beta_2$ and $\nu_t = \varepsilon_{1t} - \varepsilon_{2t}$. We'll also have $\text{Cov}[\varepsilon_{1t}, \varepsilon_{2t}] = 0$, which means that $\text{Cov}[\varepsilon_{it}, \nu_t] = \text{Var}[\varepsilon_{it}]$.⁸ These models can equivalently be expressed as

$$y_t = \min\{\mathbf{x}'_t \beta_1 + \varepsilon_{1t}, \mathbf{x}'_t \beta_2 + \varepsilon_{2t}\}. \quad (4)$$

We'll formulate a model below where the production cost is the minimum over all possible activities, and each production activity contains stochastic components (associated with inefficiency, luck, random shocks, as well as data errors and model misspecification) giving rise to an equation similar to (4).

3.1. Previous Applications of Switching Regression Models to Production Analysis

The most common application of switching regression models has been to study markets in disequilibrium. In these markets, the quantity transacted corresponds to the short side of the market and equations (2) and (3) can be used to reflect that observed quantities will sometimes lie on the demand equation (*i.e.*, regime 1) or on the supply equation (*i.e.*, regime 2); see Quandt (1982) or Maddala (1986) for a review of this literature, see Goldfeld and Quandt (1976) for a review of other uses.

Two papers concerned with production analysis use switching regression techniques, both using methods different from those proposed here. Halvorson (1985) uses switching regressions to model underground coal mining as a two-stage process. Coal is mined in the first stage and hauled to the surface in the second. Mining capacity and hauling capacity

⁸ The likelihood equation for these models when the error terms are assumed to follow a normal distribution is presented in appendix A.

both limit the output of the mining process and make the use of the ‘min’ condition in the switching regression model most appropriate. Tsurumi, Wago and Ilmakunnas (1986) are concerned with technological progress in the iron and steel Japanese industries. They use time-series data and gradual switching multivariate regression models with stochastic cross-equational constraints to model this process of change. They propose a new Bayesian procedure to detect the joint points and speed of adjustment from one regime to another.

4. A Two-Activity Switching Model

Suppose that we have a two-activity linear technology. That is, if y denotes output, and x_{ij} the amount of input j used in activity i , then we must have

$$y = \sum_{i=1}^2 \min_j \left\{ \frac{x_{ij}}{\gamma_{ij}} \right\} = \sum_{i=1}^2 \theta_i$$

where, as before, the parameter γ_{ij} measures the amount of factor j required (in addition to the pertinent amount of the other factors) to produce one unit of y when activity i is used, θ_i is the level of operation of activity i .

The cost of producing y using activity 1 is given by

$$C_1(y, \mathbf{w}) = y \sum_{j=1}^J \omega_j \gamma_{1j} = y \mathbf{w}' \Gamma_1$$

if activity 2 is used instead, we have

$$C_2(y, \mathbf{w}) = y \sum_{j=1}^J \omega_j \gamma_{2j} = y \mathbf{w}' \Gamma_2.$$

Activity 1 will be chosen by a cost-minimizing firm whenever $C_1 < C_2$, which implies

$$\mathbf{w}'(\Gamma_1 - \Gamma_2) < 0.$$

Let $\alpha = 1$ when only activity 1 is used, and $\alpha = 0$ when only activity 2 is utilized; if $\alpha \in (0, 1)$ then both activities are combined. Thus, the parameters α and $1 - \alpha$ are the intensities with which each activity is used in the production of y —*i.e.*, we have that

$$\alpha = \frac{\theta_1}{y},$$

which implies

$$\alpha = \begin{cases} 1 & \text{if } C_1 < C_2 \iff \mathbf{w}'(\Gamma_1 - \Gamma_2) < 0 \\ \text{any number in } [0, 1] & \text{if } C_1 = C_2 \iff \mathbf{w}'(\Gamma_1 - \Gamma_2) = 0 \\ 0 & \text{if } C_1 > C_2 \iff \mathbf{w}'(\Gamma_1 - \Gamma_2) > 0 \end{cases} \quad (5)$$

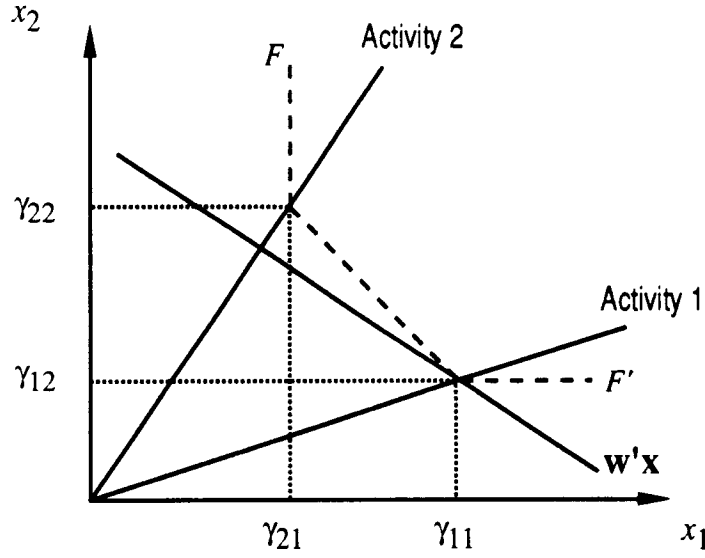


Figure 2. Example of a linear activity technology

In figure 2 we illustrate a two-input case where activity 1 would be chosen because the relative prices implied by the isocost line $\mathbf{w}'\mathbf{x}$ make it more attractive.

The conditional factor demands, x_j , for $j = 1, 2, \dots, J$, are given by

$$x_j \equiv \sum_{i=1}^2 x_{ij} = \sum_{i=1}^2 \theta_i \gamma_{ij} = y[\alpha \gamma_{1j} + (1 - \alpha) \gamma_{2j}],$$

which can be more compactly expressed in vector notation as

$$\begin{aligned} \mathbf{x} &= \theta_1 \Gamma_1 + \theta_2 \Gamma_2 \\ &= y[\alpha \Gamma_1 + (1 - \alpha) \Gamma_2]. \end{aligned} \quad (6)$$

Premultiplying both sides of equation (6) by \mathbf{w}' we obtain the conditional cost function,

$$\begin{aligned} C(y, \mathbf{w}) &= \alpha C_1(y, \mathbf{w}) + (1 - \alpha) C_2(y, \mathbf{w}) \\ &= y \mathbf{w}' [\alpha \Gamma_1 + (1 - \alpha) \Gamma_2]. \end{aligned} \quad (7)$$

We can now divide by the level of output, y , to obtain the unit average cost,

$$AC(\mathbf{w}) = \frac{C(y, \mathbf{w})}{y} = \mathbf{w}' [\alpha \Gamma_1 + (1 - \alpha) \Gamma_2].$$

Eliminating a possibly endogenous variable, output, from the right hand-side we avoid problems related to the independence of the regressors with respect to the stochastic errors.

Suppose that we observe the data vectors $(y_t, \mathbf{x}_t, \mathbf{w}_t)$ for $t = 1, \dots, T$ (where t might be different firms or different time periods). We could add an error term to each C_i in equation (7) to obtain

$$AC(\mathbf{w}_t) = \mathbf{w}'_t [\alpha_t \Gamma_1 + (1 - \alpha_t) \Gamma_2] + \alpha_t \varepsilon_{1t} + (1 - \alpha_t) \varepsilon_{2t} \quad (8)$$

where the random error must always take positive values—*i. e.*, at least the efficient quantities of each factor must be used. Equation (8) becomes an endogenous switching regression. This can be better seen when we write equation (8) as

$$\begin{aligned} AC(\mathbf{w}_t) &= \alpha_t AC_1(\mathbf{w}_t) + (1 - \alpha_t) AC_2(\mathbf{w}_t) \\ &= \min\{AC_1(\mathbf{w}_t), AC_2(\mathbf{w}_t)\} \end{aligned}$$

with

$$\begin{aligned} AC_1(\mathbf{w}_t) &= \mathbf{w}'_t \Gamma_1 + \varepsilon_{1t} \\ AC_2(\mathbf{w}_t) &= \mathbf{w}'_t \Gamma_2 + \varepsilon_{2t} \end{aligned}$$

and

$$\begin{aligned} E[\alpha_t] &= \Pr[AC_1(\mathbf{w}_t) < AC_2(\mathbf{w}_t)] \\ &= \Pr[\varepsilon_{1t} - \varepsilon_{2t} < \mathbf{w}'_t(\Gamma_2 - \Gamma_1)] \end{aligned}$$

Assuming a particular distribution for the error term in (8), we could proceed then to maximize the resulting likelihood function to obtain estimates of the relevant parameters. However, the usual maximum likelihood properties don't automatically follow since one of the regularity conditions needed to obtain them is violated; namely, the independence of the support of the random variable with respect to the values of the parameters to be estimated. Here, there is no guarantee that the error term in equation (8) will always be positive. In other words, for the error term to be always positive, the range of the left-hand side variables is bounded below by a function of the parameters of interest. This is something that the usual regularity conditions assume out. This problem was pointed out by Schmidt (1975) in relation with the procedure for estimating production frontiers proposed by Aigner and Chu (1968). Not being able to rely on general results available for maximum likelihood estimators, their properties would need to be investigated in a case-by-case basis. Greene (1980) partially solved the problem finding sufficient conditions on the error density such that maximum likelihood is consistent and asymptotically efficient (and standard errors for the estimates can be computed from the information matrix). An alternative approach is to consider the frontier itself to be random. Aigner, Lovell and Schmidt (1977) introduced such stochastic formulation for estimating frontier production function models.

5. Stochastic Cost Frontier

One possible solution to the problem posed by the non-negativity of the error term is to assume that the cost frontier itself is stochastic. Let $\varepsilon_i = \nu_i + \eta_i$, where $\nu_i > 0$. That is, equation (8) becomes

$$AC(\mathbf{w}_t) = \mathbf{w}'_t[\alpha_t \Gamma_1 + (1 - \alpha_t) \Gamma_2] + \alpha_t(\nu_{1t} + \eta_{1t}) + (1 - \alpha_t)(\nu_{2t} + \eta_{2t}) \quad (9)$$

where the ν_{it} 's are non-negative random variables related to technical and allocative inefficiency and the η_{it} 's are random variables with zero mean, $E[\eta_{it}] = 0$, which are associated with the randomness of the frontier. It follows that the average cost is bounded below by a random quantity,

$$AC(\mathbf{w}_t) \geq \mathbf{w}'_t[\alpha_t \Gamma_1 + (1 - \alpha_t) \Gamma_2] + \alpha_t \eta_{1t} + (1 - \alpha_t) \eta_{2t}$$

hence the term *stochastic cost frontier*.

5.1. Distributional Assumptions

We make now distributional assumptions on the error terms, in particular we assume that η_i is normally distributed with zero mean and variance $\sigma_{\eta_i}^2$, and ν_i is generated by a truncated-normal distribution—which, before being truncated (at zero), has mean μ_i and variance $\sigma_{\nu_i}^2$. Thus, we have

$$\nu_{it} \sim \text{iid trunc. } |N(\mu_i, \sigma_{\nu_i}^2)| \quad i = 1, 2; \forall t \quad (\text{technical inefficiency})$$

and

$$\eta_{it} \sim \text{iid } N(0, \sigma_{\eta_i}^2) \quad i = 1, 2; \forall t \quad (\text{stochastic frontier})$$

where all the covariances for which we haven't introduced any explicit notation are assumed to be zero. (See the appendix B for details on the density of a truncated-normal variable.) This specification is based in a variation of the one presented in the seminal paper by Aigner, Schmidt and Lovell (1977) due to Stevenson (1980). As pointed out by Stevenson (1980), there the inefficiency term had a mode of zero—*i.e.*, it entailed the implicit assumption that the most likely situation is one with 100% efficiency—as a result of using a truncated normal with mean zero. Here, the presence of $\mu_i > 0$ allows for some positive degree of inefficiency being the most common. In addition, the sign of the estimated μ_i can be used as a diagnosis of the model—a negative sign indicating that the model is unable to capture the characteristics of the data.

There are few alternatives to the above specification. Aigner, Amemiya and Poirier (1976) propose an error term which follows two different stochastic processes: one producing only positive disturbances and other producing only negative ones. This model has not been successful because it is hard to motivate and its likelihood equations are very complicated. Aigner, Lovell and Schmidt (1977) also present a composed-error specification which is the sum of a normal and an exponential variable; again, the mode for the inefficiency term is forced to be zero. Stevenson (1980) introduces another composed-error term which is the sum of a normal and a gamma variable; unfortunately the density function for the composed error has not been derived for the general case. Aguilar (1988) generalizes the specification in Stevenson (1980) to the case the two sources of error are not independent. He starts with a bivariate normal variable and then truncates one of the components. He derives the density of the composed-error term but fails to obtain any results on the identifiability of the model.⁹

6. Likelihood Function

From appendix B we have that the density of the composed error terms is given by

$$h_i(\varepsilon_{it}) = \frac{1}{\Phi(\mu_i/\sigma_{\nu_i}^2)\sigma_i} \phi\left(\frac{\varepsilon_{it} - \mu_i}{\sigma_i}\right) \Phi\left(\frac{\varepsilon_{it}\sigma_{\nu_i}^2 + \mu_i\sigma_{\eta_i}^2}{\sigma_{\eta_i}\sigma_{\nu_i}\sigma_i}\right) \quad i = 1, 2; \forall t \quad (10)$$

where $\sigma_i^2 = \sigma_{\nu_i}^2 + \sigma_{\eta_i}^2$; and $\phi(\cdot)$ and $\Phi(\cdot)$ are the density and the distribution function of a standard normal variable. Also, let $H_i(\varepsilon_{it}) = \int_{-\infty}^{\varepsilon_{it}} h_i(\xi)d\xi$. We are indexing the density function by i allowing for different parameters for each activity i . A more parsimonious

⁹ As a result no maximum-likelihood results can be claimed to prove desirable large-sample statistical properties for the estimators.

(and more restrictive, too) specification is possible drawing all the ν_{it} 's and η_{it} 's from the same distributions. Note, however, that if we did that, it is implicit in the switching regression specification that, at each t , we would still draw one error term for each activity. Later in the paper we present a model where not only the errors are drawn from the same distribution but only one error is drawn at each t .

The probability that AC_t is generated by activity 1 is,

$$\begin{aligned}
\lambda_t &\equiv \Pr[AC_t \text{ is generated by activity 1}] \\
&= \Pr[\mathbf{w}'_t\Gamma_1 + \varepsilon_{1t} < \mathbf{w}'_t\Gamma_2 + \varepsilon_{2t}] \\
&= \Pr[\mathbf{w}'_t(\Gamma_1 - \Gamma_2) + \varepsilon_{1t} < \varepsilon_{2t}] \\
&= \int_{-\infty}^{\infty} \int_{\mathbf{w}'_t(\Gamma_1 - \Gamma_2) + \xi_1}^{\infty} h_2(\xi_2)h_1(\xi_1)d\xi_2d\xi_1 \\
&= \int_{-\infty}^{\infty} h_1(\xi_1)[1 - H_2(\mathbf{w}'_t(\Gamma_1 - \Gamma_2) + \xi_1)]d\xi_1
\end{aligned}$$

The density of AC_t , conditional on being generated by activity 1, is proportional to

$$\int_{AC_t - \mathbf{w}'_t\Gamma_2}^{\infty} h_1(AC_t - \mathbf{w}'_t\Gamma_1)h_2(\xi)d\xi = h_1(AC_t - \mathbf{w}'_t\Gamma_1)[1 - H_2(AC_t - \mathbf{w}'_t\Gamma_2)],$$

normalizing, we obtain the conditional density

$$\frac{h_1(AC_t - \mathbf{w}'_t\Gamma_1)[1 - H_2(AC_t - \mathbf{w}'_t\Gamma_2)]}{\int_{-\infty}^{\infty} h_1(\xi - \mathbf{w}'_t\Gamma_1)[1 - H_2(\xi - \mathbf{w}'_t\Gamma_2)]d\xi}$$

but note that the denominator is simply λ_t (just change the variable of integration to $\xi_1 = \xi - \mathbf{w}'_t\Gamma_2$). Therefore, the likelihood function is given by

$$\begin{aligned}
\mathcal{L} &= \prod_{t=1}^T \left\{ \lambda_t \frac{h_1(\varepsilon_{1t})[1 - H_2(\varepsilon_{2t})]}{\lambda_t} + (1 - \lambda_t) \frac{h_2(\varepsilon_{2t})[1 - H_1(\varepsilon_{1t})]}{1 - \lambda_t} \right\} \\
&= \prod_{t=1}^T \{ h_1(\varepsilon_{1t})[1 - H_2(\varepsilon_{2t})] + h_2(\varepsilon_{2t})[1 - H_1(\varepsilon_{1t})] \}
\end{aligned} \tag{11}$$

where $\varepsilon_{it} = AC_t - \mathbf{w}'_t\Gamma_i$. Then, the log-likelihood function is

$$\begin{aligned}
\log \mathcal{L} &= \sum_{t=1}^T \log (h_1(AC_t - \mathbf{w}'_t\Gamma_1)[1 - H_2(AC_t - \mathbf{w}'_t\Gamma_2)] \\
&\quad + h_2(AC_t - \mathbf{w}'_t\Gamma_2)[1 - H_1(AC_t - \mathbf{w}'_t\Gamma_1)])
\end{aligned} \tag{12}$$

6.1. Unboundness of the Likelihood Function

A well-known problem with using maximum likelihood methods to estimate the parameters of the switching regression model is that if normality is assumed and no constraints on the error variances are imposed, then the likelihood function is unbounded for some values of the parameters [see, *e.g.*, Goldfeld and Quandt (1975), Kiefer (1978) or Quandt and Ramsey (1978)]. In our model, the disturbances aren't normally distributed but the same problem arises. Let

$$\begin{aligned} A_t &= h_1(\varepsilon_{1t})[1 - H_2(\varepsilon_{2t})] \\ B_t &= h_2(\varepsilon_{2t})[1 - H_1(\varepsilon_{1t})] \end{aligned}$$

then the likelihood function (11) can be rewritten as

$$\mathcal{L} = \prod_{t=1}^T \{A_t + B_t\}.$$

Choose $\hat{\Gamma}_1$ such that $\mathbf{w}'_1 \hat{\Gamma}_1 = AC_1$. With $\hat{\sigma}_2 \neq 0$, let $\hat{\sigma}_{\nu_1} \rightarrow 0$ and $\hat{\sigma}_{\eta_1} \rightarrow 0$ with $\hat{\sigma}_{\nu_1}/\hat{\sigma}_{\eta_1} = 1$, then $\hat{\sigma}_1 \rightarrow 0$. We have $\hat{A}_1 \rightarrow \infty$, $\hat{A}_t \rightarrow 0$ for $t \neq 1$; also $\hat{B}_t > 0$ for all t . Therefore, $\mathcal{L} \rightarrow \infty$.

This has two main consequences, one theoretical and one practical. On the theoretical side, M-class theory [Huber (1981), Amemiya (1985)] cannot be exploited to prove any large sample properties for the maximum-likelihood estimators. In a model with normal errors, Kiefer (1978) proves that a root corresponding to a local maximum of the likelihood function is both consistent and efficient. Hartley (1978) points out that there might be multiple local maxima. On the practical side, the optimization algorithm might wander returning greater and greater values of the objective function without ever converging; or it might converge to the *wrong* local maximum which the analyst has no way of distinguishing from the well-behaved root.

Goldfeld and Quandt (1975) proposed the imposition of the constraint $\sigma_2 = c\sigma_1$ to solve this problem. If we impose such constraint with known (fixed) c , then the likelihood function will be bounded for all values of the parameters. Recently, in the context of exogenous switching regressions with normal errors, Phillips (1991)¹⁰ has presented an alternative system of constraints that make the likelihood function bounded. Furthermore, conditional on the validity of the constraints, the global maximizer for the constrained likelihood function has the desirable properties of maximum likelihood estimators: it is consistent, and asymptotically normal and efficient. The constraints imposed in Phillips' method are of the form $\sigma_i \leq c\sigma_j$ with $c \in (0, 1]$, c fixed. This kind of restriction is much milder than assuming that we know the number c for which $\sigma_2 = c\sigma_1$ holds, as in Goldfeld and Quandt (1975). All that is imposed now is that neither the variance in one regime is bigger than, say, seven times the variance of the error term in the other regime. When this is imposed, Phillips (1991) proves that the global maximizer of the *constrained* likelihood function has the usual maximum likelihood desirable properties. This global maximizer can be computed with a modified EM algorithm presented in Phillips (1991).

¹⁰ Hathaway (1985) proved similar results for univariate mixture models with normal components; however in a regression context the conditional mean varies with the regressors and his results are not directly applicable.

Unfortunately, Phillips' results do not extend to our model. First, the model studied in Phillips (1991) has exogenous switching—*i.e.*, there is an exogenous probability that each observation is generated by either regime—instead of the endogenous switching that arises in our model—*i.e.*, where the 'min' condition determines which regime is in effect. Second, in our model, the disturbances won't follow a normal distribution. The end result is the much more complicated and quite unfriendly likelihood given by (11). Instead of relying in the information matrix—*i.e.*, in large sample results for MLE which we cannot yet justify—we'll propose the use of bootstrapping methods to compute their standard errors.

To solve the unboundness problem, the trick is to link the behavior of σ_2 to σ_1 so that if one goes to zero the other must go to zero as well. When that is done, in the example above we'd have $\hat{B}_t \rightarrow 0$ and the likelihood wouldn't blow up. If we restrict (σ_1, σ_2) to lie in a cone in R_+^2 —*e.g.*, by requiring $\sigma_1 \geq c\sigma_2$ and $\sigma_2 \geq c\sigma_1$ for $c \in (0, 1]$ —we cannot have one variance going to zero without driving the other to zero too. Additionally, we'll need that, for every pair (Γ_1, Γ_2) , at least one realized AC_t will be different from $\mathbf{w}'_t\Gamma_1$ and $\mathbf{w}'_t\Gamma_2$ —otherwise all the terms in \mathcal{L} would be unbounded (this can be expressed as a rank condition).

6.2. Identification

The mixture in (11) is identified whenever

$$(\mu_1, \sigma_{\eta_1}, \sigma_{\nu_1}, \Gamma_1; \mu_2, \sigma_{\eta_2}, \sigma_{\nu_2}, \Gamma_2) \neq (\mu'_1, \sigma'_{\eta_1}, \sigma'_{\nu_1}, \Gamma'_1; \mu'_2, \sigma'_{\eta_2}, \sigma'_{\nu_2}, \Gamma'_2),$$

and the second vector is not a suitable permutation of the first—*i.e.*, we don't have

$$(\mu'_1, \sigma'_{\eta_1}, \sigma'_{\nu_1}, \Gamma'_1) = (\mu_2, \sigma_{\eta_2}, \sigma_{\nu_2}, \Gamma_2)$$

and

$$(\mu'_2, \sigma'_{\eta_2}, \sigma'_{\nu_2}, \Gamma'_2) = (\mu_1, \sigma_{\eta_1}, \sigma_{\nu_1}, \Gamma_1).$$

Those permutations can be easily removed by following a labelling convention. We can index the regimes by the relative size of the coefficient of one particular parameter. For example, to remove the indeterminacy, we could impose the constraint that $\mu_1 > \mu_2$. However, in this case, this might not be a good choice since we might not want to impose the constraint that $\mu_i \neq \mu_j$. A better choice might be the input requirement of a specific factor, γ_{ij} . If we had only two inputs, then it must be the case that $\gamma_{1j} \neq \gamma_{2j}$ and the labelling convention is harmless. With more inputs that inequality must hold for, at least, two of the inputs. We can always pick an arbitrary j , and should the optimization algorithm yield close estimates for those parameters, then choose a different one. Whatever labelling condition we might use, we end up with a constrained parameter space where the likelihood is identified. This constraint—*i.e.*, labelling condition—must be taken into account when implementing an algorithm to maximize the likelihood function.

6.3. Derivatives of the Log-Likelihood Function

We present now the derivatives of the log-likelihood function (12) with respect to the parameters to be estimated:

$$(\mu_i, \sigma_{\eta_i}^2, \sigma_{\nu_i}^2, \Gamma_i), \quad \text{for } i = 1, 2.$$

These derivatives will be used to compute the MLE of these parameters. For notational convenience, let $e_{it} \equiv AC_t - \mathbf{w}'\Gamma_i$, $h_{it} \equiv h_i(e_{it})$ and $H_{it} \equiv H_i(e_{it})$. The log-likelihood function (12) can be written, then, as

$$\log \mathcal{L} = \sum_{t=1}^T \log (h_{1t}[1 - H_{2t}] + h_{2t}[1 - H_{1t}])$$

The derivatives of the log-likelihood function would be too long if collapsed into single expressions. For this reason, we express them in an iterative fashion—*i.e.*, definitions needed are shown one step below. We have

$$\begin{aligned} \frac{\partial \log \mathcal{L}}{\partial \mu_i} &= \sum_{t=1}^T \frac{\frac{\partial h_{it}}{\partial \mu_i} [1 - H_{jt}] - h_{jt} \frac{\partial H_{it}}{\partial \mu_i}}{h_{1t}[1 - H_{2t}] + h_{2t}[1 - H_{1t}]} & i, j = 1, 2; i \neq j \\ \frac{\partial \log \mathcal{L}}{\partial \sigma_{\eta_i}^2} &= \sum_{t=1}^T \frac{\frac{\partial h_{it}}{\partial \sigma_{\eta_i}^2} [1 - H_{jt}] - h_{jt} \frac{\partial H_{it}}{\partial \sigma_{\eta_i}^2}}{h_{1t}[1 - H_{2t}] + h_{2t}[1 - H_{1t}]} & i, j = 1, 2; i \neq j \\ \frac{\partial \log \mathcal{L}}{\partial \sigma_{\nu_i}^2} &= \sum_{t=1}^T \frac{\frac{\partial h_{it}}{\partial \sigma_{\nu_i}^2} [1 - H_{jt}] - h_{jt} \frac{\partial H_{it}}{\partial \sigma_{\nu_i}^2}}{h_{1t}[1 - H_{2t}] + h_{2t}[1 - H_{1t}]} & i, j = 1, 2; i \neq j \\ \frac{\partial \log \mathcal{L}}{\partial \gamma_{ik}} &= \sum_{t=1}^T \frac{\frac{\partial h_{it}}{\partial \gamma_{ik}} [1 - H_{jt}] - h_{jt} \frac{\partial H_{it}}{\partial \gamma_{ik}}}{h_{1t}[1 - H_{2t}] + h_{2t}[1 - H_{1t}]} & i, j = 1, 2; i \neq j \end{aligned}$$

where the derivatives of the cumulative density functions are

$$\frac{\partial H_i(e_{it})}{\partial \mu_i} = \int_{-\infty}^{e_{it}} \frac{\partial h_i(\xi)}{\partial \mu_i} d\xi, \quad \frac{\partial H_i(e_{it})}{\partial \sigma_{\eta_i}^2} = \int_{-\infty}^{e_{it}} \frac{\partial h_i(\xi)}{\partial \sigma_{\eta_i}^2} d\xi, \quad \frac{\partial H_i(e_{it})}{\partial \sigma_{\nu_i}^2} = \int_{-\infty}^{e_{it}} \frac{\partial h_i(\xi)}{\partial \sigma_{\nu_i}^2} d\xi.$$

and we have (see the appendix C for a detailed derivation),

$$\begin{aligned}
\frac{\partial h_i(e_{it})}{\partial \mu_i} &= -\frac{h_i(e_{it})}{\sigma_i} \left[\frac{\sigma_i \phi(\mu_i/\sigma_{\nu_i})}{\sigma_{\nu_i} \Phi(\mu_i/\sigma_{\nu_i})} - \frac{e_{it} - \mu_i}{\sigma_i} - \frac{\sigma_{\eta_i}}{\sigma_{\nu_i}} \frac{\phi\left(\frac{e_i \sigma_{\nu_i}^2 + \mu_i \sigma_{\eta_i}^2}{\sigma_{\nu_i} \sigma_{\eta_i} \sigma_i}\right)}{\Phi\left(\frac{e_i \sigma_{\nu_i}^2 + \mu_i \sigma_{\eta_i}^2}{\sigma_{\nu_i} \sigma_{\eta_i} \sigma_i}\right)} \right] \\
\frac{\partial h_i(e_{it})}{\partial \sigma_{\eta_i}^2} &= -\frac{h_i(e_{it})}{2\sigma_i^2} \left[1 - \frac{(e_{it} - \mu_i)^2}{\sigma_i^2} - \frac{\phi\left(\frac{e_i \sigma_{\nu_i}^2 + \mu_i \sigma_{\eta_i}^2}{\sigma_{\nu_i} \sigma_{\eta_i} \sigma_i}\right)}{\Phi\left(\frac{e_i \sigma_{\nu_i}^2 + \mu_i \sigma_{\eta_i}^2}{\sigma_{\nu_i} \sigma_{\eta_i} \sigma_i}\right)} \frac{\sigma_{\nu_i} (\mu_i \sigma_{\eta_i}^2 - e_{it} (\sigma_{\nu_i}^2 + 2\sigma_{\eta_i}^2))}{\sigma_{\eta_i}^3 \sigma_i} \right] \\
\frac{\partial h_i(e_{it})}{\partial \sigma_{\nu_i}^2} &= -\frac{h_i(e_{it})}{2} \left[\frac{1}{\sigma_i^2} - \frac{\mu_i \phi(\mu_i/\sigma_{\nu_i})}{\sigma_{\nu_i}^3 \Phi(\mu_i/\sigma_{\nu_i})} - \frac{(e_{it} - \mu_i)^2}{\sigma_i^4} \right. \\
&\quad \left. - \frac{\phi\left(\frac{e_i \sigma_{\nu_i}^2 + \mu_i \sigma_{\eta_i}^2}{\sigma_{\nu_i} \sigma_{\eta_i} \sigma_i}\right) \sigma_{\eta_i} (e_{it} \sigma_{\nu_i}^2 - \mu_i (\sigma_{\eta_i}^2 + 2\sigma_{\nu_i}^2))}{\Phi\left(\frac{e_i \sigma_{\nu_i}^2 + \mu_i \sigma_{\eta_i}^2}{\sigma_{\nu_i} \sigma_{\eta_i} \sigma_i}\right) \sigma_{\nu_i}^3 \sigma_i^3} \right] \\
\frac{\partial h_i(e_{it})}{\partial \gamma_{ik}} &= w_k \frac{h_i(e_{it})}{\sigma_i} \left[\frac{z - \mu_i}{\sigma_i} - \frac{\sigma_{\nu_i}}{\sigma_{\eta_i}} \frac{\phi\left(\frac{e_i \sigma_{\nu_i}^2 + \mu_i \sigma_{\eta_i}^2}{\sigma_{\nu_i} \sigma_{\eta_i} \sigma_i}\right)}{\Phi\left(\frac{e_i \sigma_{\nu_i}^2 + \mu_i \sigma_{\eta_i}^2}{\sigma_{\nu_i} \sigma_{\eta_i} \sigma_i}\right)} \right]
\end{aligned}$$

The second derivatives—which are useful for implementing some of the directional optimization algorithms—can be computed from the auxiliary derivatives in appendix B, by noting that:

$$\begin{aligned}
\frac{\partial^2 \log \mathcal{L}}{\partial \vartheta_i \partial \vartheta_j} &= \sum_{t=1}^T \frac{\frac{\partial^2 h_{it}}{\partial \vartheta_i \partial \vartheta_j} [1 - H_{jt}] - \frac{\partial h_{it}}{\partial \vartheta_i} \frac{\partial H_{jt}}{\partial \vartheta_j} - \frac{\partial h_{jt}}{\partial \vartheta_j} \frac{\partial H_{it}}{\partial \vartheta_i} - h_{jt} \frac{\partial^2 H_{it}}{\partial \vartheta_i \partial \vartheta_j}}{h_{1t}[1 - H_{2t}] + h_{2t}[1 - H_{1t}]} \\
&\quad - \sum_{t=1}^T \frac{\left[\frac{\partial h_{it}}{\partial \vartheta_i} [1 - H_{jt}] - h_{jt} \frac{\partial H_{it}}{\partial \vartheta_i} \right] \left[\frac{\partial h_{jt}}{\partial \vartheta_j} [1 - H_{it}] - h_{it} \frac{\partial H_{jt}}{\partial \vartheta_j} \right]}{[h_{1t}[1 - H_{2t}] + h_{2t}[1 - H_{1t}]]^2}
\end{aligned}$$

where ϑ_i stands for any of the parameters in $(\mu_i, \sigma_{\eta_i}^2, \sigma_{\nu_i}^2, \Gamma_i)$; the partial derivatives that appear in the expressions above—once we substitute ϑ_i and ϑ_j by the parameters of interest—are given in the appendix C. Note that

$$\frac{\partial h_{it}}{\partial \vartheta_j} = \frac{\partial^2 h_{it}}{\partial \vartheta_i \partial \vartheta_j} = \frac{\partial^2 H_{it}}{\partial \vartheta_i \partial \vartheta_j} = 0$$

unless $i = j$.

6.4. Bootstrap Estimates for the MLE Standard Errors

The bootstrap is a computer-based method for approximating the distribution of estimators or other statistics. Efron (1982) summarizes bootstrap theory; see also the survey article by Efron and Tibshirani (1986). In our case we have

$$\varepsilon_{it} \text{ iid } \sim H_i \quad \text{for } i = 1, 2; \forall t$$

where H_i is the distribution function associated with (10). Having observed the data vectors (AC_t, \mathbf{w}_t) for $t = 1, 2, \dots, T$, we estimate the parameters $(\mu_i, \sigma_{\eta_i}^2, \sigma_{\nu_i}^2, \Gamma_i)$ for $i = 1, 2$, by maximizing the likelihood (11).

The next step is to construct \hat{H}_i in order to generate the bootstrap samples. We can do that in two different ways, parametrically and nonparametrically. The parametric estimate of H_i can be obtained by simply inserting the estimated values of the parameters, $(\hat{\mu}_i, \hat{\sigma}_{\eta_i}^2, \hat{\sigma}_{\nu_i}^2)$, into H_i . This is a straightforward method and the way to proceed if we feel comfortable with the error specification fitting the data.

Alternatively, we can construct a pseudo-nonparametric estimate of H_i . One possible way of doing so is by computing the posterior probabilities that each observation AC_t was generated by either regime,

$$\begin{aligned} \hat{\pi}_{it} &= \Pr[AC_t = AC_{it} | AC_t] \\ &= \frac{h_i(AC_t - \mathbf{w}'_t \hat{\Gamma}_i | \hat{\mu}_i, \hat{\sigma}_{\eta_i}^2, \hat{\sigma}_{\nu_i}^2) [1 - H_j(AC_t - \mathbf{w}'_t \hat{\Gamma}_j | \hat{\mu}_j, \hat{\sigma}_{\eta_j}, \hat{\sigma}_{\nu_j})]}{\sum_{i=1}^2 h_i(AC_t - \mathbf{w}'_t \hat{\Gamma}_i | \hat{\mu}_i, \hat{\sigma}_{\eta_i}^2, \hat{\sigma}_{\nu_i}^2) [1 - H_j(AC_t - \mathbf{w}'_t \hat{\Gamma}_j | \hat{\mu}_j, \hat{\sigma}_{\eta_j}, \hat{\sigma}_{\nu_j})]} \end{aligned}$$

for $i, j = 1, 2; i \neq j$. We then put mass $\hat{\pi}_{it} / \sum_{s=1}^T \hat{\pi}_{is}$ at each residual

$$\hat{\varepsilon}_{it} = AC_t - \mathbf{w}'_t \hat{\Gamma}_i, \quad \text{for } i = 1, 2; t = 1, 2, \dots, T \quad (13)$$

The reason for not simply assigning mass of $1/T$ to each residual in (13) for $i = 1, 2$, is that they do not correspond to the actual errors. If we knew, for each t , that observation t was generated by regime i , then we can justifiably use $\hat{\varepsilon}_{it}$ as an estimate of ε_{it} —and we'd have two sets totalling T estimated residuals, not $2T$. However, we don't have such information and we can at best guess which regime generated each observation.¹¹ We call it *pseudo-nonparametric* since the $\hat{\pi}_{it}$'s are parametric and they're needed to construct the sample distribution function.

We are ready now to draw B bootstrap samples of size T ,¹²

$$AC_t = \min\{\mathbf{w}'_t \hat{\Gamma}_1 + e_{1t}^b, \mathbf{w}'_t \hat{\Gamma}_2 + e_{2t}^b\}$$

where

$$e_{it}^b \text{ iid } \sim \hat{H}_i \quad \text{for } i = 1, 2; t = 1, 2, \dots, T; b = 1, 2, \dots, B.$$

¹¹ We could, instead, allocate observation t to regime i whenever $\hat{\pi}_{it} > 1/2$ and give mass of $1/T_i$ to each residual in regime i where T_i is the number of observations assigned to regime i —i.e., we have $T_1 + T_2 = T$. However, this method would put too little mass in the right tail of the distribution of the disturbances—i.e., the bigger the error term the less likely that the involved regime will generate the dependent variable.

¹² It is important that the sample size of each replication be the same as the size of the original sample, T , since, after all, we are approximating the distribution of estimators based on sample size T .

For each replication b , we can compute the ‘maximum-likelihood’ estimates,

$$(\hat{\mu}_1^b, \hat{\sigma}_{\eta_1}^b, \hat{\sigma}_{\nu_1}^b, \hat{\Gamma}_1^b; \hat{\mu}_2^b, \hat{\sigma}_{\eta_2}^b, \hat{\sigma}_{\nu_2}^b, \hat{\Gamma}_2^b) \quad (14)$$

and approximate the joint distribution of the MLE estimators by a multivariate distribution function that gives mass $1/B$ to each the vectors in (14). In particular, the standard errors of the estimators can be approximated by the standard errors of the components in (14). For example, the standard error of $\hat{\mu}_1$ can be estimated by

$$\left\{ \frac{1}{B} \sum_{b=1}^B \left(\hat{\mu}_1^b - \frac{1}{B} \sum_{b=1}^B \hat{\mu}_1^b \right)^2 \right\}^{1/2}$$

We might consider the error specification given by (10) as a reasonable way to approximate the underlying distribution of the disturbance terms, allowing for the two sources of error discussed earlier. At the same time, we might not want to base our inference too heavily on it and choose the pseudo-nonparametric bootstrap. In other words, the density (10) might provide us with a way to obtain estimates of the parameters of interest but we might want to approximate the distribution of the estimators—obtained by maximizing the likelihood based on (10)—as nonparametrically as possible. There is no contradiction in proceeding this way, but just an acknowledgement that (10) might, sometimes, not be such a good approximation to the underlying process generating the data.

7. Extensions

7.1. More than two activities

We generalize now the analysis to the case of more than two activities. When we have I activities, let $\lambda_{it} = \Pr[AC_t \text{ is generated by activity } i]$. That is, we have

$$\lambda_{it} = \int_{-\infty}^{\infty} h_i(\xi_i) \prod_{l \neq i} [1 - H_l(\mathbf{w}'_t(\Gamma_i - \Gamma_l) + \xi_i)] d\xi_i$$

The density of AC_t conditional on being generated by activity i is given by

$$h_i(AC_t - \mathbf{w}'_t \Gamma_i) \prod_{l \neq i} [1 - H_l(AC_t - \mathbf{w}'_t \Gamma_l)] \frac{1}{\lambda_{it}}$$

The likelihood (11) becomes

$$\mathcal{L} = \prod_{t=1}^T \sum_{i=1}^I h_i(AC_t - \mathbf{w}'_t \Gamma_i) \prod_{l \neq i} [1 - H_l(AC_t - \mathbf{w}'_t \Gamma_l)].$$

Therefore, no theoretical complications arise when I is assumed to be greater than two, while the computational burden certainly grows with the number of activities. The constraint to make the likelihood bounded becomes

$$\max\{\sigma_i/\sigma_j\} \leq c,$$

and that for all $(\Gamma_1, \Gamma_2, \dots, \Gamma_I)$, at least for one t , the realized AC_t will be different from $\mathbf{w}'_t \Gamma_i$ for all i . The identifiability condition also generalizes easily.

7.2. Unknown Number of Activities

So far we have assumed that the number of activities, I , is known. However, it is clear than in most situations this is an heroic assumption. We'll present here an approach to testing for the number of underlying techniques.

The number of activities that can be identified is bounded above by the number of observations, but this is obviously not going to be a very useful bound in practice. However, we can perform a sequence of tests on the number of activities to determine the most likely number of them given the data. That is, we can compute tests of the form $H_0 : I = I_0$ vs $H_1 : I = I_0 + 1$ by computing the constrained MLEs,¹³ and the ratio

$$\lambda = \frac{\sup_{I=I_0+1} \mathcal{L}}{\sup_{I=I_0} \mathcal{L}}.$$

Note that I determines the number of columns in Γ so that H_0 can be considered as a test of the equality of any two columns of Γ and the pertinent elements in the error distributions—amounting to $J + 3$ restrictions. It is well known that, under regularity, $-2 \log \lambda$ is asymptotically distributed as a chi-square with r degrees of freedom where r is the number of restrictions—in this case, $r = J + 3$. Instead of appealing to the large sample distribution of $-2 \log \lambda$, its distribution can be bootstrapped. After we compute the MLE under the null and the alternative hypothesis, we can construct B bootstrap samples under each model. The test which rejects H_0 if $-2 \log \lambda$ is greater than the j th smallest of its B bootstrap replications, has size

$$\alpha = 1 - \frac{j}{B + 1}.$$

In practice, it might be computationally expensive to perform many of these tests. However, note that for each additional test we only need to generate B new bootstrap samples, not $2B$, since the B values of $\sup_{I=I_0+1} \mathcal{L}$ from the previous test can be used now for the null hypothesis.

7.3. Non-constant returns to scale

Suppose that we have

$$y = \sum_{i=1}^I \left(\min_j \left\{ \frac{x_{ij}}{\gamma_{ij}} \right\} \right)^{(1/\rho_i)}$$

so that $\rho_i < 1$ ($\rho_i > 1$) indicates the existence of increasing (decreasing) returns to scale in activity i . Then, the cost function becomes

$$C_i(y, \mathbf{w}) = \mathbf{w}' \Gamma_i y^{\rho_i}$$

and

$$AC_i(y, \mathbf{w}) = \mathbf{w}' \Gamma_i y^{\rho_i - 1}$$

The parameter vector $\rho' = (\rho_1, \rho_2, \dots, \rho_I)$ would need to be estimated now together with the rest of the parameters.

¹³ In addition to the constraints imposed by the null hypothesis, we must impose the identifiability constraint and the variance constraint discussed earlier.

8. A Different Switching Specification

There is an alternative switching specification to the model given by (4). The modification is to take error terms out of the ‘min’ condition,

$$y_t = \min\{\mathbf{x}'_t\beta_1, \mathbf{x}'_t\beta_2\} + \varepsilon_t \quad (15)$$

The model (15) is due to Ginsburgh and Zang (1975a,b; 1978). Thisler and Zang (1981) present a maximum-likelihood algorithm assuming a normal distribution for the disturbance term which is the basis for the discussion that follows.

Our activity-analysis model would result in

$$\begin{aligned} AC_t &= \min_i \{\mathbf{w}'_t\Gamma_i\} + \varepsilon_t \\ \varepsilon_t &= \nu_t + \eta_t \end{aligned} \quad (16)$$

with now the ‘min’ condition operating on deterministic linear forms. It has a straight economic interpretation since now production planning corresponds to *ex ante* plans as in Zellner, Kmenta and Drèze (1966). Equation (16) indicates that each firm would choose the activity which is expected to be the most cost effective given the factor prices. As before, the error term, ε_t , would capture the effects of noise, measurement error and exogenous shocks beyond the control of the production unit, via η_t , as well as the effects of technical inefficiency with respect to the stochastic frontier, via $\nu_t > 0$. Note that now the disturbances are not indexed by activity as before.

Assuming that

$$\nu_t \sim \text{iid trunc. } |N(\mu, \sigma_\nu^2)| \quad (\text{technical inefficiency})$$

and

$$\eta_t \sim \text{iid } N(0, \sigma_\eta^2) \quad (\text{stochastic frontier})$$

we can write the likelihood function as

$$\mathcal{L} = \prod_{t=1}^T h(AC_t - \min_i \{\mathbf{w}'_t\Gamma_i\}) \quad (17)$$

where h is ε_t 's density function derived in appendix B. The likelihood (17) is a continuous nonlinear function having discontinuous first-order partial derivatives. Having a non-smooth, non-differentiable maximand poses difficulties to the process of obtaining the maximum-likelihood estimators of the parameters of interest. Not only there are no analytical expressions for the derivatives of \mathcal{L} but the optimization methods that do not employ analytical derivatives [Powell (1964), Stewart (1967)] will fail to provide a solution since they are designed to solve differentiable problems. Thisler and Zang (1981) propose approximating the model (16) by a continuously differentiable one that can be estimated using gradient techniques.

The approximation carried out by Thisler and Zang (1981)—based on the techniques developed by Zang (1980)—consists in smoothing out the model (16) in some arbitrarily small neighbourhoods of the points where the derivatives are discontinuous leaving the original model unchanged for most values of the parameters. Define the function

$$q(x) = \min\{0, x\}, \text{ for } x \in \mathfrak{R},$$

then, $\min\{x_1, x_2\} = x_1 + q(x_2 - x_1)$ which can be applied recursively¹⁴ to express the model given by (16) as

$$AC_t = \mathbf{w}'_t \Gamma_1 + q(\mathbf{w}'_t(\Gamma_2 - \Gamma_1)) + q(\mathbf{w}'_t(\Gamma_3 - \Gamma_2)) + q(\mathbf{w}'_t(\Gamma_4 - \Gamma_3)) \\ + q(\dots + q(\mathbf{w}'_t(\Gamma_{I-2} - \Gamma_{I-1})) + q(\mathbf{w}'_t(\Gamma_I - \Gamma_{I-1})) \dots) + \varepsilon_t \quad (18)$$

The one-dimensional function $q(x)$ has a kink at $x = 0$; Zang (1980) approximates $q(x)$ by smoothing out this kink in a neighborhood of 0. That is, the function $q(x)$ is replaced by a smooth function $r(x, \delta)$ for $-\delta \leq x \leq \delta$; see figure 3.

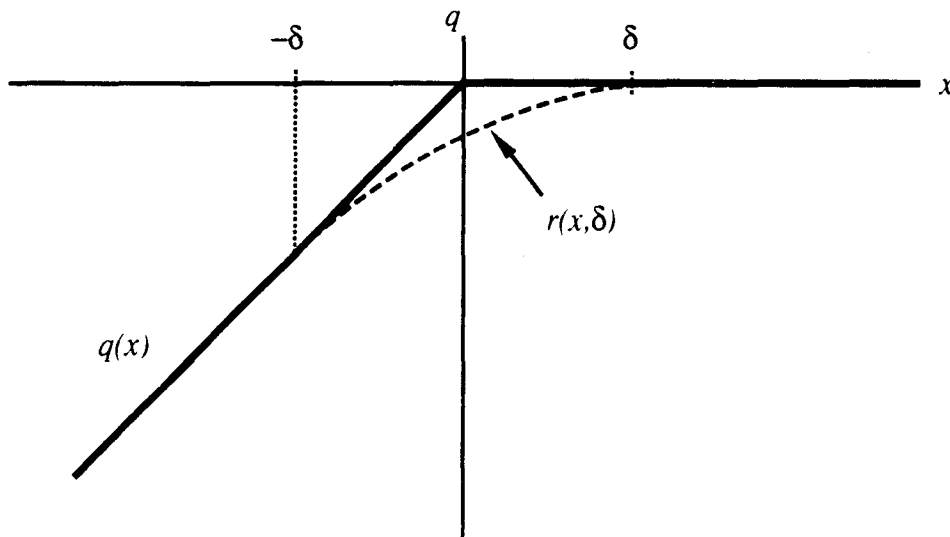


Figure 3. The function $q(x)$ and its approximation.

In particular, let's define an approximation $q_\delta(x)$ to $q(x)$ as

$$q_\delta(x) = \begin{cases} x & \text{if } x < -\delta \\ r_i(x, \delta) & \text{if } -\delta \leq x \leq \delta \\ 0 & \text{if } x > \delta \end{cases}$$

where $r_i(x, \delta)$ is a polynomial of degree $2i$ in x . Zang (1980) shows that continuous differentiability of $q_\delta(x)$ is insured up to the i th order by a polynomial of the $2i$ th order. Since we only need continuous first derivatives for ML estimation, we'll only need a polynomial of order 2. The only quadratic function that makes the approximation once continuously differentiable is

$$r_1(x, \delta) = -\frac{(\delta - x)^2}{4\delta}$$

¹⁴ Maybe an example will help to clarify this. Consider $\min\{x_1, x_2, x_3\}$, then, since $\min\{x_1, x_2\} = x_1 + q(x_2 - x_1)$ and $\min\{x_1, x_2, x_3\} = \min\{x_1, \min\{x_2, x_3\}\}$ we have that

$$\min\{x_1, x_2, x_3\} = \min\{x_1, x_2 + q(x_3 - x_2)\} = x_1 + q(x_2 - x_1 + q(x_3 - x_2)).$$

More generally, consider $\min\{x_1, x_2, \dots, x_n\}$, and let $\Delta_i = x_{i+1} - x_i$. Then

$$\min\{x_1, x_2, \dots, x_n\} = x_1 + q(\Delta_1 + q(\Delta_2 + q(\Delta_3 + \dots + q(\Delta_{n-1}))) \dots).$$

therefore, our approximated $q(x)$ is given by

$$q_\delta(x) = \begin{cases} x & \text{if } x < -\delta \\ -(\delta - x)^2/4\delta & \text{if } -\delta \leq x \leq \delta \\ 0 & \text{if } x > \delta \end{cases} \quad (19)$$

We can substitute our model (18) by a smooth continuously differentiable approximation which is identical to the original model except in an arbitrarily small neighborhood of the T^{J-1} possible kinks. The approximation is given by

$$AC_t \approx \mathbf{w}'_t \Gamma_1 + q_\delta(\mathbf{w}'_t(\Gamma_2 - \Gamma_1)) + q_\delta(\mathbf{w}'_t(\Gamma_3 - \Gamma_2)) + q_\delta(\mathbf{w}'_t(\Gamma_4 - \Gamma_3)) \\ + q_\delta(\dots + q_\delta(\mathbf{w}'_t(\Gamma_{I-2} - \Gamma_{I-1})) + q_\delta(\mathbf{w}'_t(\Gamma_I - \Gamma_{I-1}))) \dots + \varepsilon_t \quad (20)$$

and $q_\delta(x)$ is given by (19).

Inserting this approximation into the likelihood function (17), we obtain

$$\log \mathcal{L}_\delta = \sum_{t=1}^T \log h(AC_t - Q_\delta(\mathbf{w}_t, \Gamma)) \quad (21)$$

where, for convenience, we write

$$Q_\delta(\mathbf{w}_t, \Gamma) = \mathbf{w}'_t \Gamma_1 + q_\delta(\mathbf{w}'_t(\Gamma_2 - \Gamma_1)) + q_\delta(\mathbf{w}'_t(\Gamma_3 - \Gamma_2)) + q_\delta(\mathbf{w}'_t(\Gamma_4 - \Gamma_3)) \\ + q_\delta(\dots + q_\delta(\mathbf{w}'_t(\Gamma_{I-2} - \Gamma_{I-1})) + q_\delta(\mathbf{w}'_t(\Gamma_I - \Gamma_{I-1}))) \dots \quad (22)$$

8.1. Derivatives of the Log-Likelihood Function

Define $e_t = AC_t - Q_\delta(\mathbf{w}_t, \Gamma)$. Using the results in the appendix C, we obtain the likelihood equations,

$$\begin{aligned} \frac{\partial \log \mathcal{L}_\delta}{\partial \mu} &= -\frac{T}{\sigma_\nu} \frac{\phi(\mu/\sigma_\nu)}{\Phi(\mu/\sigma_\nu)} + \frac{1}{\sigma} \sum_{t=1}^T \left[\frac{e_t - \mu}{\sigma} + \frac{\sigma_\eta}{\sigma_\nu} \frac{\phi\left(\frac{e_t \sigma_\nu^2 + \mu \sigma_\eta^2}{\sigma_\eta \sigma_\nu \sigma}\right)}{\Phi\left(\frac{e_t \sigma_\nu^2 + \mu \sigma_\eta^2}{\sigma_\eta \sigma_\nu \sigma}\right)} \right] \\ \frac{\partial \log \mathcal{L}_\delta}{\partial \sigma_\eta^2} &= -\frac{T}{2\sigma^2} + \frac{1}{2\sigma^2} \sum_{t=1}^T \left[\frac{(e_t - \mu)^2}{\sigma^2} + \frac{\phi\left(\frac{e_t \sigma_\nu^2 + \mu \sigma_\eta^2}{\sigma_\eta \sigma_\nu \sigma}\right)}{\Phi\left(\frac{e_t \sigma_\nu^2 + \mu \sigma_\eta^2}{\sigma_\eta \sigma_\nu \sigma}\right)} \frac{\sigma_\nu (\mu \sigma_\eta - e_t (\sigma_\nu^2 + 2\sigma_\eta^2))}{\sigma_\eta^3 \sigma^3} \right] \\ \frac{\partial \log \mathcal{L}_\delta}{\partial \sigma_\nu^2} &= -\frac{T}{2} \left[\frac{1}{\sigma^2} - \frac{\phi(\mu/\sigma_\nu) \mu}{\Phi(\mu/\sigma_\nu) \sigma_\nu^3} \right] \\ &\quad + \frac{1}{2\sigma^3} \sum_{t=1}^T \left[\frac{(e_t - \mu)^2}{\sigma} + \frac{\phi\left(\frac{e_t \sigma_\nu^2 + \mu \sigma_\eta^2}{\sigma_\eta \sigma_\nu \sigma}\right)}{\Phi\left(\frac{e_t \sigma_\nu^2 + \mu \sigma_\eta^2}{\sigma_\eta \sigma_\nu \sigma}\right)} \frac{\sigma_\eta (e_t \sigma_\nu^2 - \mu (\sigma_\eta^2 + 2\sigma_\nu^2))}{\sigma_\nu^3} \right] \\ \frac{\partial \log \mathcal{L}_\delta}{\partial \gamma_{ij}} &= \frac{1}{\sigma} \sum_{t=1}^T \left[\frac{e_t - \mu}{\sigma^2} - \frac{\sigma_\nu}{\sigma_\eta} \frac{\phi\left(\frac{e_t \sigma_\nu^2 + \mu \sigma_\eta^2}{\sigma_\eta \sigma_\nu \sigma}\right)}{\Phi\left(\frac{e_t \sigma_\nu^2 + \mu \sigma_\eta^2}{\sigma_\eta \sigma_\nu \sigma}\right)} \right] \frac{\partial Q_\delta}{\partial \gamma_{ij}} \end{aligned}$$

where $\partial Q_\delta / \partial \gamma_{ij}$ can be recursively computed from (22). The second derivatives can be computed from the results in the appendix C by noting that

$$\frac{\partial \log \mathcal{L}_\delta}{\partial \vartheta} = \sum_{t=1}^T D_\vartheta(e_t)$$

where ϑ represents any of the parameters, μ , σ_η^2 , σ_ν^2 , or γ_{ij} . That is, if we want $\partial^2 \log \mathcal{L}_\delta / \partial \mu \partial \sigma_\eta^2$ we have that

$$\frac{\partial^2 \log \mathcal{L}_\delta}{\partial \mu \partial \sigma_\eta^2} = \sum_{t=1}^T \frac{\partial D_\mu(e_t)}{\partial \sigma_\eta^2}$$

and the expression for $\partial D_\mu(e_t) / \partial \sigma_\eta^2$ can be found in the appendix C.

8.2. Maximization of the Log-Likelihood Function

The maximization of the log-likelihood equation can be done by solving the likelihood equations (*e.g.*, by quasi-Newton methods) for an arbitrarily small δ .¹⁵ The scheme consists of solving a sequence of minimization programs corresponding to a sequence $\{\delta^{(k)}\} \rightarrow 0$. In practice, we don't need to get δ arbitrarily close to 0, we just need a value such that $\mathcal{L} = \mathcal{L}_\delta$ at the solution to the likelihood equations. In other words, the approximation given by (20) is just a local approximation in the neighborhood of the kink points. The probability of getting data points precisely at the kinks is zero. Making the neighborhoods small enough, the approximated model and the original model will coincide when evaluated at the sample points and we end up estimating the original model exactly. The approximation scheme can be interpreted as an optimal classification strategy.

An iterative algorithm [adapted from Thisler and Zang (1981)] is presented below. Note that step (k).2 is just a way to check that the values of the original likelihood, \mathcal{L} , and the values of the likelihood of the approximated model, \mathcal{L}_δ , coincide—if that is so, the approximation Q_δ is no longer operating and the exact model is being estimated.

[Step (k).1] Solve the likelihood equations with $\delta = \delta^{(k)}$. (A suitable initial value for δ^1 is some value less than the average AC_t .) Let the estimated values of the parameters be $(\mu^{(k)}, \sigma_\eta^{(k)}, \sigma_\nu^{(k)}; \Gamma_1^{(k)}, \Gamma_2^{(k)}, \dots, \Gamma_J^{(k)})$.

[Step (k).2] Let

$$\Delta^{(k)} = \min_t \{B_t^{(k)} - C_t^{(k)}\}$$

where

$$B_t^{(k)} = \min_j \{\mathbf{w}'_t \Gamma_j^{(k)}\}$$

and

$$C_t^{(k)} = \min_j \{\mathbf{w}'_t \Gamma_j^{(k)} : \mathbf{w}'_t \Gamma_j^{(k)} > B_t^{(k)}\}$$

if $\Delta^{(k)} \geq 2\delta^{(k)}$ then $(\mu^{(k)}, \sigma_\eta^{(k)}, \sigma_\nu^{(k)}; \Gamma_1^{(k)}, \Gamma_2^{(k)}, \dots, \Gamma_J^{(k)})$ is an optimal solution to the original problem, then STOP. If, on the contrary, we have $\Delta^{(k)} < 2\delta^{(k)}$ go to step (k).3.

¹⁵ If we wanted to use the Hessian of the log-likelihood function in the optimization algorithm, we'd use a fourth order polynomial in place of $r_1(x, \delta)$ —in particular, we'd use $r_2(x, \delta) = -(x - \delta)^4 / 16\delta^3$.

[Step (k).3] If $\Delta^{(k)} = 0$ then set $\delta^{(k+1)} = \delta^{(k)}/d$. Otherwise, let

$$\delta^{(k+1)} = \frac{\Delta^{(k)}}{d}$$

with $d \geq 2$, where d is some fixed parameter. If $\delta^{(k+1)} < \epsilon$ where ϵ is a tolerance parameter then STOP and declare $(\mu^{(k)}, \sigma_\eta^{(k)}, \sigma_\nu^{(k)}; \Gamma_1^{(k)}, \Gamma_2^{(k)}, \dots, \Gamma_J^{(k)})$ an optimal solution. Otherwise go to step (k + 1).1.

Let

$$(\mu^*, \sigma_\eta^*, \sigma_\nu^*; \Gamma_1^*, \Gamma_2^*, \dots, \Gamma_J^*)$$

be a solution to the optimization algorithm (k).1–(k).3 finishing in a step (k*).2; then it is a local maximum of the original likelihood equation (17). The approximation is no longer active and the function q_δ is replaced back by q . If the algorithm has stopped in a step (k*).3, then we are at a kink for some t and reducing δ beyond the tolerance parameter can result in numerical problems. We can declare the solution a maximum of (17) or use direct search methods like the downhill simplex method [see Nelder and Mead (1965)] to maximize the likelihood.

In the absence of perfect multicollinearity, the model is locally identified around the maximum likelihood estimate obtained by the previous procedure provided we have ‘enough’ observations in each regime [Lubrano (1985)]. That is, if we have J inputs (*i.e.*, J parameters in each Γ_i) then we must have at least $J + 1$ observations assigned to each regime. Thus, $T/(J + 1)$ gives us an upper bound for the maximum number of activities that we might hope to identify.

The same bootstrapping methods presented for the other switching specification can be used with this model with the advantage of becoming simpler—since only one error is drawn at each t , and the common distribution implies that less parameters are involved.

9. Comparison of the Two Types of Models

We have presented two alternative switching specifications; the classical switching regression model:

$$AC_t = \min_i \{ \mathbf{w}'_t \Gamma_i + \varepsilon_{it} \} \quad (23)$$

and the model introduced in the previous section:

$$AC_t = \min_i \{ \mathbf{w}'_t \Gamma_i \} + \varepsilon_t \quad (24)$$

We want to discuss which model fits best in our activity-analysis context. The key to answer this question is to distinguish between the stochastic effects which are known to the firm and part of its decisions and the stochastic effects which are unknown to the firm. The disturbance term allows for variation of the cost frontier across observations; the symmetric part, η , representing noise, measurement error, and exogenous shocks beyond the control of the production unit—*e.g.*, climate, topography, luck—and errors of measurement in AC . The one-sided part, ν , captures deviations due to technical inefficiencies within the firm.

Aigner, Lovell and Schmidt (1977) argue that ν represents things under the control of the firms while η represents events not under the control of the firm (but not observable by the econometrician). Under those assumptions, the correct model would become

$$AC_t = \min_i \{ \mathbf{w}'_t \Gamma_i + \nu_{it} \} + \eta_t \quad (25)$$

The model specified by equation (25) is a combination of the models (23) and (24) which we studied in this paper. Start with model (23) and assume that we draw all the η_{it} 's from a common distribution, and, for each t , only one η_t is drawn—*i.e.*, as opposed to drawing one η_{it} for each activity from a common distribution—then, we arrive at model (25).

Schmidt (1985) argues against assuming that technical inefficiency is totally unforeseen [Schmidt (1985), pp. 294–295, 304–305]. Cases can be made about what to include inside the 'min' condition and what to leave out but, unfortunately, this model is intractable because of its complexity and a compromise must be made. In each context, the analyst must choose which model is more appropriate to carry out the analysis and the choice will depend, on part, on the specific conditions and objectives of the study. We must choose between models (23) and (24).

Aigner, Lovell and Schmidt (1977) find that, in their applications, the symmetric component, η , dominates the one sided component, ν —*i.e.*, they note that σ_η^2 is much bigger than σ_ν^2 . However, Lee and Tyler (1978) and Kopp and Smith (1978) present examples where this is not the case.

10. Concluding Remarks

The empirical analysis of production, especially frontier analysis, has been a very active field of research in recent years [see, *e.g.*, Lewin and Lovell (1990)]. The literature is divided in two major approaches: linear programming approach (nonparametric and nonstatistical), and the econometric approach (parametric and statistical). In this paper we have developed methods that combine the advantages of both methodologies: functional flexibility of the first approach and the statistical nature of the second approach. This combination comes at the cost of some deal of analytical complexity.

Directions for future research are the derivation of large sample properties for the estimators presented here, as well as some Monte Carlo analysis to assess their small sample behavior. Apart from the analytical complexity involved, we do not foresee special difficulties in establishing the usual consistency, efficiency and asymptotic normality of the maximum likelihood estimators. On the other hand, preliminary Monte Carlo analysis suggests that possibly a reparametrization of the models presented in this paper must be done in order to improve the numerical performance of the optimization algorithms. In particular, instead of estimating σ_ν^2 and σ_η^2 separately it might be better—from a numerical point of view—to estimate the ratio $\sigma_\eta^2/\sigma_\nu^2$ and the sum $\sigma^2 = \sigma_\eta^2 + \sigma_\nu^2$.

APPENDIX A

Likelihood Function of Normal Endogenous Switching Regression Models

Assume that the error terms in (4) are independently and normally distributed with zero means and variances σ_i^2 , for $i = 1, 2$. Then, the error density, $f_i(\cdot)$, and distribution, $F_i(\cdot)$, functions, are given by

$$\begin{aligned} f_i(\varepsilon_i) &= \frac{1}{\sigma_i} \phi(\varepsilon_i/\sigma_i), \\ F_i(\varepsilon_i) &= \Phi(\varepsilon_i/\sigma_i); \end{aligned}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the density and distribution functions of a normal variable.

The probability that observation t belongs to regime 1 is given by

$$\begin{aligned} \lambda_t &= \Pr[y_t = y_{1t}] \\ &= \Pr[\mathbf{x}'_t \beta_1 + \varepsilon_{1t} < \mathbf{x}'_t \beta_2 + \varepsilon_{2t}] \\ &= \Pr[\varepsilon_{1t} + \mathbf{x}'_t (\beta_1 - \beta_2) < \varepsilon_{2t}] \\ &= \int_{-\infty}^{\infty} \left\{ \int_{\xi_1 + \mathbf{x}'_t (\beta_1 - \beta_2)}^{\infty} f_2(\xi_2) d\xi_2 \right\} f_1(\xi_1) d\xi_1 \\ &= \int_{-\infty}^{\infty} \left[1 - \Phi\left(\frac{\xi_1 + \mathbf{x}'_t (\beta_1 - \beta_2)}{\sigma_2}\right) \right] f_1(\xi_1) d\xi_1 \\ &= \int_{-\infty}^{\infty} \left[1 - \Phi\left(\frac{\xi \sigma_1 + \mathbf{x}'_t (\beta_1 - \beta_2)}{\sigma_2}\right) \right] \phi(\xi) d\xi \\ &= \Phi\left(\frac{\mathbf{x}'_t (\beta_2 - \beta_1)}{\sigma_1^2 + \sigma_2^2}\right). \end{aligned} \tag{26}$$

(The last equality follows from the fact that $\varepsilon_1 + \varepsilon_2$ is a normal variable with zero mean and variance $\sigma_1^2 + \sigma_2^2$.) Define, for $i = 1, 2$,

$$g_i(y) = f_i(y - \mathbf{x}' \beta_i), \text{ and } G_i(y) = \int_{-\infty}^y f_i(\xi) d\xi.$$

Conditional on y_t belonging to regime 1, its density is given by

$$h(y_t | y_t = y_{1t}) = \frac{\int_{y_t}^{\infty} g_1(y_t) g_2(y_2) dy_2}{\int_{-\infty}^{\infty} \int_y^{\infty} g_1(y) g_2(y_2) dy_2 dy} = \frac{g_1(y_t) [1 - G_2(y_t)]}{\lambda_t}.$$

Similarly, $h(y_t | y_t = y_{2t}) = g_2(y_t) [1 - G_1(y_t)] / (1 - \lambda_t)$. Then, the (unconditional) likelihood function is given by

$$\begin{aligned} \mathcal{L} &= \prod_{t=1}^T \{ \Pr[y_t = y_{1t}] h(y_t | y_t = y_{1t}) + \Pr[y_t = y_{2t}] h(y_t | y_t = y_{2t}) \} \\ &= \prod_{t=1}^T \left\{ \lambda_t \frac{g_1(y_t) [1 - G_2(y_t)]}{\lambda_t} + (1 - \lambda_t) \frac{g_1(y_t) [1 - G_1(y_t)]}{1 - \lambda_t} \right\} \\ &= \prod_{t=1}^T \{ g_1(y_t) [1 - G_2(y_t)] + g_1(y_t) [1 - G_1(y_t)] \}. \end{aligned}$$

APPENDIX B

Distribution of the Sum of a Normal and Truncated Normal Variables

First, let's introduce some convenient notation. Let $\phi(\cdot)$ denote the standard normal density,

$$\phi(s) = \frac{1}{\sqrt{2\pi}} e^{-s^2/2} 1_{(-\infty, \infty)}(s)$$

where $1_A(x)$ is an indicator function which takes the value of 1 when $x \in A$ and 0 otherwise. Also, let $\Phi(\cdot)$ denote the standard normal distribution function,

$$\Phi(t) = \int_{-\infty}^t \phi(s) ds.$$

Let X follow a normal distribution with zero mean and variance σ_x^2 ; *i.e.*, let its density function be

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp \left\{ -\frac{1}{2} \left(\frac{x}{\sigma_x} \right)^2 \right\} 1_{(-\infty, \infty)}(x)$$

and let $F(\cdot)$ be its cumulative density function, $F(x) = \int_{-\infty}^x f(\xi) d\xi$. Also, let Y follow a truncated-normal distribution, $|N(\mu, \sigma_y)|$; *i.e.*, its density is given by

$$g(y) = \frac{1}{\Phi(\mu/\sigma_y)\sqrt{2\pi}\sigma_y} \exp \left\{ -\frac{1}{2} \left(\frac{y - \mu}{\sigma_y} \right)^2 \right\} 1_{(0, \infty)}(y)$$

and let $G(\cdot)$ be its cumulative density function, $G(x) = \int_0^x g(\xi) d\xi$.

The probability that $X + Y \leq z$, $H(z)$, is given by

$$\begin{aligned} H(z) &= \int_0^\infty \int_{-\infty}^{z-y} f(x)g(y) dx dy \\ &= \int_0^\infty F(z-y)g(y) dy \end{aligned}$$

differentiating with respect to z , we obtain the density function for $Z = X + Y$,

$$\begin{aligned} h(z) &= \int_0^\infty f(z-y)g(y) dy \\ &= \frac{1}{\Phi(\mu/\sigma_y)2\pi\sigma_x\sigma_y} \int_0^\infty \exp \left\{ -\frac{1}{2} \left(\left(\frac{z-y}{\sigma_x} \right)^2 + \left(\frac{y-\mu}{\sigma_y} \right)^2 \right) \right\} dy \end{aligned} \quad (27)$$

We shall establish that

$$h(z) = \frac{1}{\Phi(\mu/\sigma_y)\sigma} \phi \left(\frac{z-\mu}{\sigma} \right) \Phi \left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma} \right) \quad (28)$$

where $\sigma^2 = \sigma_x^2 + \sigma_y^2$.

Let's first look at the exponent in (27),

$$\begin{aligned} \left(\frac{z-y}{\sigma_x}\right)^2 + \left(\frac{y-\mu}{\sigma_y}\right)^2 &= \left(\frac{y\sigma^2 - (z\sigma_y^2 + \mu\sigma_x^2)}{\sigma_x\sigma_y\sigma}\right)^2 - \left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right)^2 + \frac{z^2}{\sigma_x^2} + \frac{\mu^2}{\sigma_y^2} \\ &= \left(\frac{y\sigma^2 - (z\sigma_y^2 + \mu\sigma_x^2)}{\sigma_x\sigma_y\sigma}\right)^2 + \left(\frac{z-\mu}{\sigma}\right)^2 \end{aligned}$$

Next, we'll change the variable of integration, let

$$\xi = \frac{y\sigma^2 - (z\sigma_y^2 + \mu\sigma_x^2)}{\sigma_x\sigma_y\sigma}$$

which implies that $d\xi = (\sigma/\sigma_x\sigma_y)dy$; also, the lower limit of the integral must be changed,

$$y = 0 \quad \Rightarrow \quad \xi = -\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}$$

Therefore, we have

$$\begin{aligned} h(z) &= \frac{1}{\Phi(\mu/\sigma_y)2\pi\sigma_x\sigma_y} \int_0^\infty \exp\left\{-\frac{1}{2}\left(\left(\frac{z-y}{\sigma_x}\right)^2 + \left(\frac{y-\mu}{\sigma_y}\right)^2\right)\right\} dy \\ &= \frac{1}{\Phi(\mu/\sigma_y)\sigma} \left[\frac{e^{-(z-\mu)^2/2\sigma^2}}{\sqrt{2\pi}}\right] \left[\int_{-\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}}^\infty \frac{e^{-\xi^2/2}}{\sqrt{2\pi}} d\xi\right] \\ &= \frac{1}{\Phi(\mu/\sigma_y)\sigma} \phi\left(\frac{z-\mu}{\sigma}\right) \Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right) \end{aligned}$$

as desired.

APPENDIX C

Auxiliary Derivatives

In this appendix we compute the derivatives of the density developed in the previous appendix, $h(z)$, with respect to the parameters that enter into the econometric model. These results are used in the paper to compute the likelihood equations.

First, we compute the derivatives of the standard normal density and distribution function,

$$\frac{\partial \phi(z)}{\partial z} = -\frac{z}{\sqrt{2\pi}} e^{-z^2/2} = -z\phi(z)$$

and $\partial \Phi(z)/\partial z = \phi(z)$.

We compute, now, the derivatives of the error density and distribution function with respect to its argument (we'll need this derivative when taking $\partial \mathcal{L}/\partial \Gamma_i$, for example),

$$\begin{aligned} \frac{\partial h(z)}{\partial z} &= -\frac{\phi\left(\frac{z-\mu}{\sigma}\right)}{\Phi(\mu/\sigma_y)\sigma} \left[\frac{z-\mu}{\sigma^2} \Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right) - \frac{\sigma_y}{\sigma_x\sigma} \phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right) \right] \\ &= -\frac{h(z)}{\sigma} \left[\frac{z-\mu}{\sigma} - \frac{\sigma_y}{\sigma_x} \frac{\phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right)}{\Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right)} \right] \end{aligned}$$

and $\partial H(z)/\partial z = h(z)$.

Next, we compute the derivatives of the error density with respect to the density parameters. We have

$$\begin{aligned} \frac{\partial h(z)}{\partial \mu} &= -\frac{\phi(\mu/\sigma_y)}{\sigma\Phi(\mu/\sigma_y)^2\sigma_y} \phi\left(\frac{z-\mu}{\sigma}\right) \Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right) \\ &\quad + \frac{1}{\sigma\Phi(\mu/\sigma_y)} \frac{z-\mu}{\sigma^2} \phi\left(\frac{z-\mu}{\sigma}\right) \Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right) \\ &\quad + \frac{1}{\sigma\Phi(\mu/\sigma_y)} \phi\left(\frac{z-\mu}{\sigma}\right) \frac{\sigma_x}{\sigma_y\sigma} \phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right) \\ &= -\frac{h(z)}{\sigma} \left[\frac{\sigma}{\sigma_y} \frac{\phi(\mu/\sigma_y)}{\Phi(\mu/\sigma_y)} - \frac{z-\mu}{\sigma} - \frac{\sigma_x}{\sigma_y} \frac{\phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right)}{\Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right)} \right] \end{aligned}$$

Also, noting that $\sigma^2 = \sigma_x^2 + \sigma_y^2$, we get

$$\begin{aligned}
\frac{\partial h(z)}{\partial \sigma_x^2} &= -\frac{1}{2\Phi(\mu/\sigma_y)\sigma^3} \phi\left(\frac{z-\mu}{\sigma}\right) \Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right) \\
&+ \frac{1}{\Phi(\mu/\sigma_y)\sigma} \frac{(z-\mu)^2}{2\sigma^4} \phi\left(\frac{z-\mu}{\sigma}\right) \Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right) \\
&+ \frac{1}{\Phi(\mu/\sigma_y)\sigma} \phi\left(\frac{z-\mu}{\sigma}\right) \phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right) \frac{\mu\sigma_x^2\sigma_y^2 - z\sigma_y^2(\sigma_y^2 + 2\sigma_x^2)}{2\sigma_x^3\sigma_y\sigma^3} \\
&= -\frac{h(z)}{2\sigma^2} \left[1 - \frac{(z-\mu)^2}{\sigma^2} - \frac{\phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right)}{\Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right)} \frac{\sigma_y(\mu\sigma_x^2 - z(\sigma_y^2 + 2\sigma_x^2))}{\sigma_x^3\sigma} \right]
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial h(z)}{\partial \sigma_y^2} &= -\frac{1}{2\Phi(\mu/\sigma_y)\sigma^3} + \frac{\mu\phi(\mu/\sigma_y)}{2\Phi(\mu/\sigma_y)^2\sigma_y^3} \phi\left(\frac{z-\mu}{\sigma}\right) \Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right) \\
&+ \frac{1}{\Phi(\mu/\sigma_y)\sigma} \frac{(z-\mu)^2}{2\sigma^4} \phi\left(\frac{z-\mu}{\sigma}\right) \Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right) \\
&+ \frac{1}{\Phi(\mu/\sigma_y)\sigma} \phi\left(\frac{z-\mu}{\sigma}\right) \phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right) \frac{z\sigma_x^2\sigma_y^2 - \mu\sigma_x^2(\sigma_x^2 + 2\sigma_y^2)}{2\sigma_x\sigma_y^3\sigma^3} \\
&= -\frac{h(z)}{2} \left[\frac{1}{\sigma^2} - \frac{\phi(\mu/\sigma_y)\mu}{\Phi(\mu/\sigma_y)\sigma_y^3} - \frac{(z-\mu)^2}{\sigma^4} \right. \\
&\quad \left. - \frac{\phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right)}{\Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right)} \frac{\sigma_x(z\sigma_y^2 - \mu(\sigma_x^2 + 2\sigma_y^2))}{\sigma_y^3\sigma^3} \right]
\end{aligned}$$

We'll introduce some convenient notation before we write expressions for the second derivatives. Let

$$D_z(z) = \frac{\partial \log h(z)}{\partial z} = -\frac{1}{\sigma} \left[\frac{z-\mu}{\sigma} - \frac{\sigma_y}{\sigma_x} \frac{\phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right)}{\Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right)} \right]$$

and

$$D_\mu(z) = \frac{\partial \log h(z)}{\partial \mu} = -\frac{1}{\sigma} \left[\frac{\sigma}{\sigma_y} \frac{\phi(\mu/\sigma_y)}{\Phi(\mu/\sigma_y)} - \frac{z-\mu}{\sigma} - \frac{\sigma_x}{\sigma_y} \frac{\phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right)}{\Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right)} \right]$$

$$D_{\sigma_x^2}(z) = \frac{\partial \log h(z)}{\partial \sigma_x^2} = -\frac{1}{2\sigma^2} \left[1 - \frac{(z-\mu)^2}{\sigma^2} - \frac{\phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right)}{\Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right)} \frac{\mu\sigma_x^2\sigma_y - z\sigma_y(\sigma_y^2 + 2\sigma_x^2)}{\sigma_x^3\sigma} \right]$$

$$D_{\sigma_y^2}(z) = \frac{\partial \log h(z)}{\partial \sigma_y^2} = -\frac{1}{2} \left[\frac{1}{\sigma^2} - \frac{\phi(\mu/\sigma_y)\mu}{\Phi(\mu/\sigma_y)\sigma_y^3} - \frac{(z-\mu)^2}{\sigma^4} - \frac{\phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right) z\sigma_x\sigma_y^2 - \mu\sigma_x(\sigma_x^2 + 2\sigma_y^2)}{\Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right) \sigma_y^3\sigma^3} \right]$$

Then, we can write:

$$\begin{aligned} \frac{\partial h(z)}{\partial z} &= D_z(z)h(z) & \frac{\partial h(z)}{\partial \sigma_x^2} &= D_{\sigma_x^2}(z)h(z) \\ \frac{\partial h(z)}{\partial \mu} &= D_\mu(z)h(z) & \frac{\partial h(z)}{\partial \sigma_y^2} &= D_{\sigma_y^2}(z)h(z) \end{aligned}$$

and the second-order derivatives can be written as:

$$\begin{aligned} \frac{\partial^2 h(z)}{\partial z^2} &= \frac{\partial D_z(z)}{\partial z} h(z) + D_z(z) \frac{\partial h(z)}{\partial z} = \left(\frac{\partial D_z(z)}{\partial z} + D_z^2 \right) h(z) \\ \frac{\partial^2 h(z)}{\partial \mu^2} &= \frac{\partial D_\mu(z)}{\partial \mu} h(z) + D_\mu(z) \frac{\partial h(z)}{\partial \mu} = \left(\frac{\partial D_\mu(z)}{\partial \mu} + D_\mu^2 \right) h(z) \\ \frac{\partial^2 h(z)}{\partial \sigma_x^4} &= \frac{\partial D_{\sigma_x^2}(z)}{\partial \sigma_x^2} h(z) + D_{\sigma_x^2}(z) \frac{\partial h(z)}{\partial \sigma_x^2} = \left(\frac{\partial D_{\sigma_x^2}(z)}{\partial \sigma_x^2} + D_{\sigma_x^2}^2 \right) h(z) \\ \frac{\partial^2 h(z)}{\partial \sigma_y^4} &= \frac{\partial D_{\sigma_y^2}(z)}{\partial \sigma_y^2} h(z) + D_{\sigma_y^2}(z) \frac{\partial h(z)}{\partial \sigma_y^2} = \left(\frac{\partial D_{\sigma_y^2}(z)}{\partial \sigma_y^2} + D_{\sigma_y^2}^2 \right) h(z) \end{aligned}$$

and

$$\begin{aligned} \frac{\partial^2 h(z)}{\partial z \partial \mu} &= \frac{\partial D_z(z)}{\partial \mu} h(z) + D_z(z) \frac{\partial h(z)}{\partial \mu} = \left(\frac{\partial D_z(z)}{\partial \mu} + D_z(z) D_\mu(z) \right) h(z) \\ &= \frac{\partial D_\mu(z)}{\partial z} h(z) + D_\mu(z) \frac{\partial h(z)}{\partial z} = \left(\frac{\partial D_\mu(z)}{\partial z} + D_z(z) D_\mu(z) \right) h(z) \\ \frac{\partial^2 h(z)}{\partial z \partial \sigma_x^2} &= \frac{\partial D_z(z)}{\partial \sigma_x^2} h(z) + D_z(z) \frac{\partial h(z)}{\partial \sigma_x^2} = \left(\frac{\partial D_z(z)}{\partial \sigma_x^2} + D_z(z) D_{\sigma_x^2}(z) \right) h(z) \\ &= \frac{\partial D_{\sigma_x^2}(z)}{\partial z} h(z) + D_{\sigma_x^2}(z) \frac{\partial h(z)}{\partial z} = \left(\frac{\partial D_{\sigma_x^2}(z)}{\partial z} + D_z(z) D_{\sigma_x^2}(z) \right) h(z) \\ \frac{\partial^2 h(z)}{\partial z \partial \sigma_y^2} &= \frac{\partial D_z(z)}{\partial \sigma_y^2} h(z) + D_z(z) \frac{\partial h(z)}{\partial \sigma_y^2} = \left(\frac{\partial D_z(z)}{\partial \sigma_y^2} + D_z(z) D_{\sigma_y^2}(z) \right) h(z) \\ &= \frac{\partial D_{\sigma_y^2}(z)}{\partial z} h(z) + D_{\sigma_y^2}(z) \frac{\partial h(z)}{\partial z} = \left(\frac{\partial D_{\sigma_y^2}(z)}{\partial z} + D_z(z) D_{\sigma_y^2}(z) \right) h(z) \\ \frac{\partial^2 h(z)}{\partial \mu \partial \sigma_x^2} &= \frac{\partial D_\mu(z)}{\partial \sigma_x^2} h(z) + D_\mu(z) \frac{\partial h(z)}{\partial \sigma_x^2} = \left(\frac{\partial D_\mu(z)}{\partial \sigma_x^2} + D_\mu(z) D_{\sigma_x^2}(z) \right) h(z) \\ &= \frac{\partial D_{\sigma_x^2}(z)}{\partial \mu} h(z) + D_{\sigma_x^2}(z) \frac{\partial h(z)}{\partial \mu} = \left(\frac{\partial D_{\sigma_x^2}(z)}{\partial \mu} + D_\mu(z) D_{\sigma_x^2}(z) \right) h(z) \end{aligned}$$

$$\begin{aligned}\frac{\partial^2 h(z)}{\partial \mu \partial \sigma_y^2} &= \frac{\partial D_\mu(z)}{\partial \sigma_y^2} h(z) + D_\mu(z) \frac{\partial h(z)}{\partial \sigma_y^2} = \left(\frac{\partial D_\mu(z)}{\partial \sigma_y^2} + D_\mu(z) D_{\sigma_y^2}(z) \right) h(z) \\ &= \frac{\partial D_{\sigma_y^2}(z)}{\partial \mu} h(z) + D_{\sigma_y^2}(z) \frac{\partial h(z)}{\partial \mu} = \left(\frac{\partial D_{\sigma_y^2}(z)}{\partial \mu} + D_\mu(z) D_{\sigma_y^2}(z) \right) h(z)\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 h(z)}{\partial \sigma_x^2 \partial \sigma_y^2} &= \frac{\partial D_{\sigma_x^2}(z)}{\partial \sigma_y^2} h(z) + D_{\sigma_x^2}(z) \frac{\partial h(z)}{\partial \sigma_y^2} = \left(\frac{\partial D_{\sigma_x^2}(z)}{\partial \sigma_y^2} + D_{\sigma_x^2}(z) D_{\sigma_y^2}(z) \right) h(z) \\ &= \frac{\partial D_{\sigma_y^2}(z)}{\partial \sigma_x^2} h(z) + D_{\sigma_y^2}(z) \frac{\partial h(z)}{\partial \sigma_x^2} = \left(\frac{\partial D_{\sigma_y^2}(z)}{\partial \sigma_x^2} + D_{\sigma_x^2}(z) D_{\sigma_y^2}(z) \right) h(z)\end{aligned}$$

Finally, noting that

$$\frac{\partial \phi(\psi(z))}{\partial z \Phi(\psi(z))} = -\frac{\phi(z)[\Phi(\psi(z))\psi'(z) + \phi(\psi(z))\psi'(z)]}{\Phi(\psi(z))^2}$$

we have that

$$\frac{\partial D_z(z)}{\partial z} = -\frac{1}{\sigma^2} - \frac{\sigma_y^2}{\sigma_x^2 \sigma^2} \phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x \sigma_y \sigma}\right) \frac{\Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x \sigma_y \sigma}\right) \left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x \sigma_y \sigma}\right) + \phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x \sigma_y \sigma}\right)}{\Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x \sigma_y \sigma}\right)^2}$$

$$\frac{\partial D_z(z)}{\partial \mu} = \frac{1}{\sigma^2} - \frac{1}{\sigma^2} \phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x \sigma_y \sigma}\right) \frac{\Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x \sigma_y \sigma}\right) \left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x \sigma_y \sigma}\right) + \phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x \sigma_y \sigma}\right)}{\Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x \sigma_y \sigma}\right)^2}$$

$$\begin{aligned}\frac{\partial D_z(z)}{\partial \sigma_x^2} &= \frac{z - \mu}{\sigma^4} - \frac{\sigma_y(\sigma_x^2 + \sigma^2)}{2\sigma_x^3 \sigma^3} \frac{\phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x \sigma_y \sigma}\right)}{\Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x \sigma_y \sigma}\right)} \\ &\quad - \phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x \sigma_y \sigma}\right) \frac{\Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x \sigma_y \sigma}\right) \left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x \sigma_y \sigma}\right) + \phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x \sigma_y \sigma}\right)}{\Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x \sigma_y \sigma}\right)^2} \\ &\quad \times \frac{\sigma_y^2(\mu\sigma_x^2 - z(\sigma_y^2 + 2\sigma_x^2))}{2\sigma_x^4 \sigma^4}\end{aligned}$$

$$\begin{aligned}\frac{\partial D_z(z)}{\partial \sigma_y^2} &= \frac{z - \mu}{\sigma^4} + \frac{\sigma_x}{2\sigma_y \sigma^3} \frac{\phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x \sigma_y \sigma}\right)}{\Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x \sigma_y \sigma}\right)} \\ &\quad + \phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x \sigma_y \sigma}\right) \frac{\Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x \sigma_y \sigma}\right) \left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x \sigma_y \sigma}\right) + \phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x \sigma_y \sigma}\right)}{\Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x \sigma_y \sigma}\right)^2} \\ &\quad \times \frac{z\sigma_y^2 - \mu(\sigma_x^2 + 2\sigma_y^2)}{2\sigma_y^2 \sigma^4}\end{aligned}$$

$$\begin{aligned} \frac{\partial D_\mu(z)}{\partial \mu} &= -\frac{1}{\sigma^2} + \frac{1}{\sigma_y^2} \phi(\mu/\sigma_y) \frac{\Phi(\mu/\sigma_y)(\mu/\sigma_y) + \phi(\mu/\sigma_y)}{\Phi(\mu/\sigma_y)^2} \\ &\quad - \phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right) \frac{\Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right) \left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right) + \phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right)}{\Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right)^2} \frac{\sigma_x^2}{\sigma_y^2\sigma^2} \end{aligned}$$

$$\begin{aligned} \frac{\partial D_\mu(z)}{\partial \sigma_x^2} &= -\frac{z - \mu}{\sigma^4} + \frac{\sigma_y}{2\sigma_x\sigma^3} \frac{\phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right)}{\Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right)} \\ &\quad - \phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right) \frac{\Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right) \left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right) + \phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right)}{\Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right)^2} \\ &\quad \times \frac{\mu\sigma_x^2 - z(\sigma_y^2 + 2\sigma_x^2)}{2\sigma_x^2\sigma^4} \end{aligned}$$

$$\begin{aligned} \frac{\partial D_\mu(z)}{\partial \sigma_y^2} &= -\frac{z - \mu}{\sigma^4} - \frac{\sigma_x(\sigma_y^2 + \sigma^2)}{2\sigma_y^3\sigma^3} \frac{\phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right)}{\Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right)} \\ &\quad + \frac{1}{2\sigma_y^3} \frac{\phi(\mu/\sigma_y)}{\Phi(\mu/\sigma_y)} + \frac{\mu}{2\sigma_y^4} \phi(\mu/\sigma_y) \frac{\Phi(\mu/\sigma_y)(\mu/\sigma_y) + \phi(\mu/\sigma_y)}{\Phi(\mu/\sigma_y)^2} \\ &\quad - \phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right) \frac{\Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right) \left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right) + \phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right)}{\Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right)^2} \\ &\quad \times \frac{\sigma_x^2(z\sigma_y^2 - \mu(\sigma_x^2 + 2\sigma_y^2))}{2\sigma_y^4\sigma^4} \end{aligned}$$

$$\begin{aligned} \frac{\partial D_{\sigma_x^2}(z)}{\partial \sigma_x^2} &= \frac{1}{2\sigma^4} - \frac{(z - \mu)^2}{\sigma^6} \\ &\quad + \frac{\phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right)}{\Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right)} \frac{\sigma_x^2\sigma_y^2(8z\sigma^2 - \mu(\sigma_y^2 + 4\sigma_x^2)) + 3z\sigma^6}{4\sigma_x^5\sigma_y\sigma^5} \\ &\quad - \phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right) \frac{\Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right) \left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right) + \phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right)}{\Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right)^2} \\ &\quad \times \frac{\sigma_y^2(\mu\sigma_x^2 - z(\sigma_y^2 + 2\sigma_x^2))^2}{4\sigma_x^6\sigma^6} \end{aligned}$$

$$\begin{aligned}
\frac{\partial D_{\sigma_x^2}(z)}{\partial \sigma_y^2} &= \frac{1}{2\sigma^4} - \frac{(z-\mu)^2}{\sigma^6} \\
&+ \frac{\phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right) \mu\sigma_x^2 - 2\mu\sigma_y^2 - 2z\sigma_x^2 + z\sigma_y^2}{\Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right) 4\sigma_x\sigma_y\sigma^5} \\
&- \phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right) \frac{\Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right) \left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right) + \phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right)}{\Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right)^2} \\
&\times \frac{(\mu\sigma_x^2 - z(\sigma_y^2 + 2\sigma_x^2))(z\sigma_y^2 - \mu(\sigma_x^2 + 2\sigma_y^2))}{4\sigma_x^2\sigma_y^2\sigma^6}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial D_{\sigma_y^2}(z)}{\partial \sigma_x^2} &= \frac{1}{\sigma^4} - \frac{(z-\mu)^2}{2\sigma^6} - \frac{3\mu}{4\sigma_y^5} \frac{\phi(\mu\sigma_y)}{\Phi(\mu\sigma_y)} \\
&+ \phi(\mu/s_y) \frac{\Phi(\mu/s_y)(\mu/s_y) + \phi(\mu/s_y) \mu^2}{\Phi(\mu/s_y)^2} \frac{\mu^2}{4\sigma_y^6} \\
&+ \frac{\phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right) 3\mu\sigma_x^6 + 8\mu\sigma_x^2\sigma_y^2\sigma^2 - z\sigma_x^2\sigma_y^2(\sigma_x^2 + 4\sigma_y^2)}{\Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right) 4\sigma_x\sigma_y^5\sigma^5} \\
&- \phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right) \frac{\Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right) \left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right) + \phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right)}{\Phi\left(\frac{z\sigma_y^2 + \mu\sigma_x^2}{\sigma_x\sigma_y\sigma}\right)^2} \\
&\times \frac{\sigma_x^2(z\sigma_y^2 - \mu(\sigma_x^2 + 2\sigma_y^2))^2}{4\sigma_y^6\sigma^6}
\end{aligned}$$

References

- Aguilar, R. (1988): *Efficiency in Production: Theory and Application on Kenyan Smallholders*. Göteborg: Ekonomiska studier utgivna av Nationalekonomiska institutionen vid Göteborgs universitet.
- 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: Maximum Likelihood Estimation of the Parameters of a Discontinuous Density Function," *International Economic Review*, **17**:2, 377–96.
- Aigner, D., C.A.K. Lovell and P. Schmidt (1977): "Formulation and Estimation of Stochastic Frontier Production Function Models," *Journal of Econometrics*, **6**, 21–37.
- Amemiya, T. (1985): *Advanced Econometrics*. Cambridge, MA: Harvard University Press.
- Banker, R.D., A. Charnes and W.W. Cooper (1984): "Some Models for Estimating Technical and Returns to Scale Inefficiencies in Data Envelopment Analysis," *Management Science*, **30**:9, 1078–92.
- Banker, R.D. and A. Maindiratta (1987): "Maximum Likelihood Estimation of Monotone Concave production Frontiers," unpublished manuscript, December.
- Banker, R.D. (1988): "Maximum Likelihood, Consistency and Data Envelopment Analysis: A Statistical Foundation," unpublished manuscript, Carnegie-Mellon University, July.
- Bauer, P.W. (1990): "Recent Econometric Developments in Frontier Estimation," *Journal of Econometrics*, **46**:1, 39–56.
- Chamberlain, G. (1980): "Analysis of Covariance with Qualitative Data," *Review of Economic Studies*, **47**, 225–238.
- Charnes, A., W.W. Cooper and E. Rhodes (1978): "Measuring the Efficiency of Decision Making Units," *European Journal of Operational Research*, **2**:6, 429–44.
- Efron, B. (1982): *The Jackknife, the Bootstrap and Other Resampling Plans*. Philadelphia: SIAM.
- Efron, B. and R. Tibshirani (1986): "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy," *Statistical Science*, **1**:1, 54–77.
- Farrell, M.J. (1957): "The Measurement of Productive Efficiency," *Journal of the Royal Statistical Society, Series A*, **120**:3, 253–81.
- Färe, R., S. Grosskopf and C.A.K. Lovell (1985): *The Measurement of Efficiency of Production*. Boston: Kluwer-Nijhoff Publishing.

- Førsund, F.R., C.A.K. Lovell and P. Schmidt (1980): "A Survey of Frontier Production Functions and of their Relationship to Efficiency Measurement," *Journal of Econometrics*, **13**:1, 5–25.
- Ginsburgh, V.A. and I. Zang (1975a): "Price Taking or Price Making Behavior: An Alternative to Full Cost Price Functions," Cowles Foundation Discussion Paper no. 403.
- Ginsburgh, V.A. and I. Zang (1975b): "Price Taking or Price Making Behavior in Export Pricing," CORE Discussion Paper no. 7805.
- Goldfeld, S. and R. Quandt (1973): "The Estimation of Structural Shifts by Switching Regressions," *Annals of Economic and Social Measurement*, **2**:4, 475–485.
- Goldfeld, S. and R. Quandt (1975): "Estimation in a Disequilibrium Model and the Value of Information," *Journal of Econometrics*, **3**:3, 325–348.
- Greene, W.H. (1980): "Maximum Likelihood Estimation of Econometric Frontier Functions," *Journal of Econometrics*, **13**:1, 101–115.
- Halvorson, A.L. (1985): "Switching regression estimates of a sequential production process: the case of underground coal mining," *The Review of Economics and Statistics*, **67**, 161–5.
- Hartley, M.J. (1978): "Comment on 'Estimating Mixtures of Normal Distributions and Switching Regressions'," *Journal of the American Statistical Association*, **73**:364, 738–741.
- Hathaway, R.J. (1985): "A Constrained Formulation of Maximum-Likelihood Estimates for Normal Mixture Distributions," *Annals of Statistics*, **13**, 795–800.
- Huber, P.J. (1981): *Robust Statistics*. New York: John Wiley & Sons.
- Kiefer, N.M. (1978): "Discrete Parameter Variation: Efficient Estimation of a Switching Regression Model," *Econometrica*, **46**:2, 427–434.
- Kiefer, N.M. (1980): "A Note on Switching Regressions and Logistic Discrimination." *Econometrica*, **48**:4, 1065–1069.
- Koopmans, T. (1951): "Analysis of Production as an Efficient Combination of Activities," in *Activity Analysis of Production and Allocation*, ed. T. Koopmans. New York: John Wiley & Sons.
- Kopp, R. and V.K. Smith (1978): "The Characteristics of Frontier Production Estimates for Steam Generating Plants: An Econometric Analysis," *Southern Economic Journal*, **47**, 1049–1059.
- Land, C.K., C.A.K. Lovell and S. Thore (1988): "Chance-Constrained Efficiency Analysis," Paper presented at the NSF Conference on Parametric and Nonparametric Approaches to Frontier Analysis, September 30–October 1, 1988; University of North Carolina at Chapel Hill.

- Lee, L.-F. and W.G. Tyler (1978): "A Stochastic Frontier Production Function and Average Efficiency: An Empirical Analysis," *Journal of Econometrics*, **7**, 385–390.
- Lewin, A.Y. and C.A.K. Lovell (1990): "Frontier Analysis: Parametric and Nonparametric Approaches," *Journal of Econometrics*, **46**:1/2, .
- Lubrano, M. (1985): "Bayesian analysis of switching regression models," *Journal of Econometrics*, **29**, 69–95.
- Maddala, G.S. and F.D. Nelson (1974): "Maximum Likelihood Methods for Models of Markets in Disequilibrium," *Econometrica*, **42**:6, 1013–1030.
- Maddala, G.S. (1986): "Disequilibrium, Self-selection, and Switching Models," in *Handbook of Econometrics*, vol. III, ed. Z. Griliches and M. D. Intriligator. New York: North-Holland.
- Nelder, J.A. and R. Mead (1965): "A Simplex Method for Function Minimization," *Computer Journal*, **7**, 308–313.
- Phillips, R.F. (1991): "A Constrained Maximum-Likelihood Approach to Estimating Switching Regressions," *Journal of Econometrics*, **48**, 241–262.
- Pole, A.M. and A.F.M. Smith (1985): "A Bayesian analysis of some threshold switching models," *Journal of Econometrics*, **29**, 97–119.
- Powell, M.J.D. (1964): "An Efficient Method for Finding the Minimum of Function of Several Variables without Calculating Derivatives," *Computer Journal*, **7**, 155–162.
- Quandt, R.E. and J.B. Ramsey (1978): "Estimating Mixtures of Normal Distributions and Switching Regressions," *Journal of the American Statistical Society*, **73**:364, 730–738.
- Quandt, R.E. (1982): "Econometric Disequilibrium Models," *Econometric Reviews*, **1**:1, 1–63.
- Schmidt, Peter (1975): "On The Statistical Estimation of Parametric Production Functions," *The Review of Economic and Statistics*, **58**:2, 238–129.
- Schmidt, Peter (1985): "Frontier Production Functions," *Econometric Reviews*, **4**:2, 289–328.
- Seiford, L.M. and R.M. Thrall (1990): "Recent Econometric Developments in DEA," *Journal of Econometrics*, **46**, 7–38.
- Shephard, R.W. (1970): *Theory of Cost and Production Functions*. Princeton: Princeton University Press.
- Stevenson, R.E. (1980): "Likelihood Functions for Generalized Stochastic Frontier Estimation," *Journal of Econometrics*, **13**:1, 58–66.
- Stewart, G.W. (1967): "A Modification of Davidon's Minimization Method to Accept Difference Approximation of Derivatives," *Journal of the Association of Computing Machinery*, **14**, 72–83.

- Thisler, A. and I. Zang (1981): "A New Maximum Likelihood Algorithm for Piecewise Regression," *Journal of the American Statistical Association*, **76**:376, 980–987.
- Tsurumi, H., H. Wago and P. Ilmakunnas (1986): "Gradual Switching Multivariate Regression Models with Stochastic Cross-Equational Constraints and an Application to the KLEM Translog Production Model," *Journal of Econometrics*, **31**, 235–53.
- Varian, H.R. (1985): "Nonparametric Tests of Optimizing Behavior with Measurement Error," *Journal of Econometrics*, **30**, 445–458.
- Zang, I. (1978): "A Smoothing-Out Technique for Min-Max Optimization," *Mathematical Programming*, **19**, 61–77.
- Zellner, A., J. Kmenta and J. Drèze (1966): "Specification and Estimation of Cobb-Douglas Production Function Models," *Econometrica*, **34**, 784–795.