Empirical Economics (2003) 28:581–598

EMPIRICAL
ECONOMICS

Estimating a mixture of stochastic frontier regression models via the em algorithm: A multiproduct cost function application

Steven B. Caudill*

Department of Economics, Auburn University, Auburn University, AL 36849
(e-mail: SCAUDILL@BUSINESS.AUBURN.EDU)

First version received: 3/13/01/Final version received: 6/17/02

Abstract. Researchers have become increasingly interested in estimating mixtures of stochastic frontiers. Mester (1993), Caudill (1993), and Polachek and Yoon (1987), for example, estimate stochastic frontier models for different regimes, assuming sample separation information is given. Building on earlier work by Lee and Porter (1984), Douglas, Conway, and Ferrier (1995) estimate a stochastic frontier switching regression model in the presence of noisy sample separation information. The purpose of this paper is to extend earlier work by estimating a mixture of stochastic frontiers assuming *no* sample separation information. This case is more likely to occur in practice than even noisy sample separation information. In order to estimate a mixture of stochastic frontiers with no sample separation information, an EM algorithm to obtain maximum likelihood estimates is developed. The algorithm is used to estimate a mixture of stochastic (cost) frontiers using data on U.S. savings and loans for the years 1986, 1987, and 1988. Statistical evidence is found supporting the existence of a mixture of stochastic frontiers.

Key words: Mixture model, Stochastic frontier, efficiency

JEL: C24, C81, D24

1. Introduction

For many years economists interested in examining firm efficiency have estimated regression models with composite errors. These regression models are

* I am grateful to Ram Acharya, Janice Caudill, and especially James R. Barth for several helpful comments on an earlier version of the paper. During the revision process I benefitted greatly from the suggestions of the Associate Editor and three anonymous referees.

called stochastic frontier models and they have been estimated in numerous studies (for an overview of the applications of frontier methods in the financial services industry, see Berger and Mester (1997)). Researchers have recently become interested in estimating mixtures of stochastic frontiers. Mester (1993), Caudill (1993), and Polachek and Yoon (1987), for example, estimate stochastic frontier models for different regimes, assuming sample separation information is given. Extending earlier work by Lee and Porter (1984), Douglas, Conway, and Ferrier (1995) also estimate a stochastic frontier switching regression model in which sample separation information is available, but may be incorrect. This paper extends earlier work by Douglas, Conway, and Ferrier by developing an expectations maximization, or EM, algorithm to estimate a mixture of stochastic frontiers in the presence of *no* sample separation information. This case is more likely to occur in practice than noisy sample separation information but presents more computational difficulties.

The estimation of a mixture of stochastic frontier regressions represents a breakthrough in the measurement of firm efficiency. Earlier work by Beard, Caudill, and Gropper (1991, 1997) estimated a mixture of two cost functions. BCG justify the estimation of a mixture model by arguing that firms in an industry may use different, but unobservable, technologies. If the different technologies in use could be observed, firms could be separated into groups based on the underlying technology prior to estimation and separate cost functions could be estimated for each. If the technology each firm employs is not observed, that is no sample separation information is available, cost functions associated with different regimes can still be estimated with a mixture model. In the estimation of mixture models, firms are probabilistically separated into groups and separate cost functions are estimated for each without any information about which firms use which technology. The estimation of a mixture model also permits statistical testing of the null hypothesis of a single cost regime.

Furthermore, if two or more underlying technologies are present and a single cost function is estimated, a specification error results that can lead to misleading conclusions about efficiency rankings. The method presented here avoids problems caused by this specification error by extending the earlier work of BCG to the estimation of a mixture of stochastic frontiers.

In order to estimate a mixture of stochastic frontiers, this paper develops an EM algorithm to obtain maximum likelihood estimates. The algorithm has its origins in the work of Dempster, Laird, and Rubin (1977) who adapted the EM algorithm for use in incomplete data problems. The EM algorithm presented here is directly based on the work of Huang (1984) and Hartley (1978), who developed EM algorithms for the maximum likelihood estimation of stochastic frontiers and mixtures of normal regressions, respectively.[1]

Why estimate a mixture model using the EM algorithm? There are several well-known algorithms available to estimate model parameters by maximum likelihood, but most, like the algorithm of Berndt, Hall, Hall, and Hausman (1974), have problems estimating mixture models due to singularities in the likelihood surface.[2] Until the work of Hartley (1978), the estimation of mix-

[1] An excellent reference to many applications of the EM algorithm, along with details, is provided by McLachlan and Krishnan (1997). For an important early application of the EM algorithm in economics see Stewart (1983).

[2] For a discussion of these problems see, for example, Quandt (1988), section 2.4, pp. 35 to 40.

ture models by maximum likelihood was questionable. The advantages and disadvantages of the EM algorithm when compared to Newton-Raphson and the method of scoring are discussed by Titterington, Smith, and Makov (1985, pp. 88–89). These authors list the strengths of the EM algorithm as being: 1) simplicity of application, and 2) guaranteed monotonic convergence to, at least, a local maximum. The weaknesses are: 1) slower convergence than other methods, 2) no estimator of covariance matrix is provided, and 3) there is no guarantee of convergence to a global maximum.

For the estimation problem considered here, the advantages of the EM algorithm clearly outweigh the disadvantages. The major benefit of EM is programming simplicity. As for the disadvantages, 1) EM's slower convergence rate is not particularly troublesome because computing time is not a concern, 2) variance estimates are obtained by using a single iteration of the algorithm of Berndt, Hall, Hall, and Hausman (1974) on the converged EM values, and 3) none of the available derivative-based methods is guaranteed to converge to a global maximum.[3] For the estimation of a mixture of stochastic frontiers, the EM algorithm has many virtues.

The paper is organized as follows. First, the mixture of stochastic frontiers model is developed along with a discussion of estimation by maximum likelihood using the EM algorithm. Next, stochastic frontier models, mixture models, and a mixture of stochastic frontiers model are estimated using data on U.S. savings and loans for the years 1986, 1987, and 1988. The statistical evidence supports the presence of a mixture of stochastic frontiers.

2. A mixture of stochastic frontiers

Maximum likelihood estimation via the EM algorithm for a mixture of stochastic frontiers requires that the two EM algorithms for estimating stochastic frontiers and normal mixtures be combined into a single EM algorithm. In the following sections the EM algorithms for the estimation of both stochastic frontiers and mixtures models are detailed. Then, these results are combined into a single EM algorithm for the estimation of a mixture of stochastic frontiers by maximum likelihood.

The Stochastic Frontier Model. The composite error model examined here is called a stochastic frontier regression model (see Aigner, Lovell, and Schmidt (1977), Battese and Corra (1977), or Meeusen and van den Broeck (1977)). The composite error in this regression model is the sum (or difference) of a symmetric error and a one-sided error. In this model, the symmetric error is assumed to be normally distributed and the one-sided error is assumed to be half-normally distributed.[4] These two error components are assumed to be independent. Following Huang (1984), the stochastic frontier (cost) regression

[3] There are other non-derivative based algorithms that are guaranteed to converge to a global maximum, for example, the simulated annealing approach of Goffe, Ferrier, and Rogers (1994).

[4] Stevenson (1980) generalized the usual half-normal assumption for the one-sided error by considering a normal distribution truncated at points other than zero. Lee (1983) assumed the one-sided error term came from a four parameter Pearson family of distributions. Beckers and Hammond (1987) and Greene (1990) assumed a gamma distribution. Recently, some of the specifications of the inefficiency error have been distribution-free, see Berger (1993).

model can be written as

$$Y_i = Z_i + v_i$$

where

$$\begin{aligned} Z_i &= X_i\beta + w_i \\ \Rightarrow Y_i &= X_i\beta + w_i + v_i \end{aligned} \tag{1}$$

where Y_i is observable cost, Z_i, is the unobservable frontier cost, and v_i is the half-normally distributed random variable derived from a normally distributed random variable with mean zero and variance, σ_v^2. This unobservable frontier cost, Z_i, is related to several independent variables, X_i, a vector of unknown parameters to be estimated, β, and a random error term, w_i, assumed to be normally distributed with mean zero and variance, σ_w^2.

Huang (1984) provides the complete data loglikelihood function, based on the joint density of Y and Z, which is given by

$$f(Y_i, Z_i) = \frac{1}{\pi\sigma_v\sigma_w}\exp\left\{-\frac{1}{2}\left(\frac{Z_i - X_i\beta}{\sigma_w}\right)^2 - \frac{1}{2}\left(\frac{Y_i - Z_i}{\sigma_v}\right)^2\right\}. \tag{2}$$

First order conditions can be obtained by differentiating (2), first with respect to β, which yields

$$\frac{\partial E\ln L}{\partial\beta} = \sum_{i=1}^{n}(E(Z_i|\,Y_i) - X_i\beta)X_i' = 0. \tag{3}$$

The solution of this equation is the familiar least squares regression estimator with X_i as the vector of independent variables and $E(Z_i|\,Y_i)$ as the dependent variable. The remaining derivatives yield

$$\begin{aligned} \sigma_v^2 &= \frac{1}{n}\sum_{i=1}^{n}(Y_i^2 - 2E(Z_i|\,Y_i)Y_i + E(Z_i^2|\,Y_i)) \\ \sigma_w^2 &= \frac{1}{n}\sum_{i=1}^{n}(E(Z_i^2|\,Y_i) - 2E(Z_i|\,Y_i)X_i\beta + (X_i\beta)^2). \end{aligned} \tag{4}$$

To evaluate the expressions above, moments of the conditional density of Z given Y are needed. The conditional density of Z_i given Y_i is shown by Huang and Jondrow, Lovell, Materov, and Schmidt (1982) to be $N(\mu_i^*, \sigma_*^2)$ truncated from above at Y_i, where

$$\mu_i^* = \sigma^{-2}[\sigma_w^2 Y_i + \sigma_v^2 X_i\beta] \quad \text{and} \quad \sigma_*^2 = \sigma_w^2\sigma_v^2/\sigma^2 \tag{5}$$

where

$$\sigma^2 = \sigma_w^2 + \sigma_v^2.$$

The EM algorithm requires the first two moments of the resulting truncated normal density. Letting $Y_i^* = (Y_i - \mu)/\sigma$, these moments, with the truncation in this case being from above, are given by Maddala (1983a) as

$$E(Z_i|Y_i) = E(Z_i \,|\, Z_i \le Y_i) = \mu_i^* - \sigma_* \frac{f_i^*}{F_i^*}$$

$$E(Z_i^2|Y_i) = E(Z_i^2 \,|\, Z_i \le Y_i) = \sigma_*^2 \left[1 - Y_i^* \frac{f_i^*}{F_i^*} - \left(\frac{f_i^*}{F_i^*}\right)^2 \left(\mu_i^* - \sigma_* \frac{f_i^*}{F_i^*}\right)^2 \right] \tag{6}$$

where f_i^* and F_i^* are the standard normal density and distribution functions, respectively, evaluated at Y_i^*. These moments can be substituted into (3) and (4) and evaluated iteratively until convergence.

A Mixture of Normal Regressions. The estimation of a mixture of normal regression models is a problem that has recently received much attention.[5] Early work on mixture models or switching regressions is provided by Quandt (1972), Quandt and Ramsey (1978), and Goldfeld and Quandt (1976). These models are closely related to disequilibrium models. For important early work in this area, see Fair and Jaffee (1972), Maddala and Nelson (1974), and Maddala (1983b).

An early application of mixture models in the production area is given by Aigner, Amemiya, and Poirer (1976). These authors estimate a regression model in which the error term is a mixture of a positive and negative half-normal. Recently, mixtures of statistical cost functions with normal errors have been estimated by Beard, Caudill, and Gropper (1991) and Beard, Caudill, and Gropper (1997). These authors argue that firms in an industry may use different technologies of production and, as a consequence, may operate on different cost functions. If no information is available about which firms use which technologies (no sample separation information is available), BCG argue that the estimation of a mixture model is appropriate.

Hartley (1978) and Quandt (1988) provide the details of the estimation of a mixture of normal regressions by maximum likelihood using the EM algorithm. For the case of a mixture of two normal regressions (or switching regressions), consider

$$\begin{aligned} Y_i &= X_i\beta_1 + \varepsilon_{1i} \quad \text{with probability } \theta \\ Y_i &= X_i\beta_2 + \varepsilon_{2i} \quad \text{with probability } 1-\theta, \end{aligned} \tag{7}$$

where ε_{1i} and ε_{2i} are mutually independent, iid normal with zero means and variances σ_1^2 and σ_2^2, respectively. Following Hartley (1978), the typical complete-data density function is given by

$$\begin{aligned} f(d_{i1}, Y_i) = {} & \left\{ \frac{\theta}{\sqrt{2\pi}\sigma_1} \exp\left\{ -\frac{(Y_i - X_i\beta_1)^2}{2\sigma_1^2} \right\} \right\}^{d_{i1}} \\ & \times \left\{ \frac{1-\theta}{\sqrt{2\pi}\sigma_2} \exp\left\{ -\frac{(Y_i - X_i\beta_2)^2}{2\sigma_2^2} \right\} \right\}^{1-d_{i1}}. \end{aligned} \tag{8}$$

[5] For overviews of mixture models, see McLachlan and Basford (1988), Everitt and Hand (1981), and Titterington, Smith, and Makov (1985).

These densities comprise the logarithm of the complete-data likelihood function that is given by

$$\ln L = \sum_{i=1}^{n} \{d_{i1}(\ln\theta + \ln f_{i1}) + (1 - d_{i1})(\ln(1-\theta) + \ln f_{i2})\}, \tag{9}$$

where f_{i1} and f_{i2} are the respective normal density functions.

In the E step of the EM algorithm, the expected value of the loglikelihood is needed, which requires replacing d in (9) by its expectation given the data. This expectation is given by $E(d_{i1}|Y_i) = (1)[(P(d_{i1} = 1 \mid Y_i)] + (0)[P(d_{i1} = 0 \mid Y_i)] = P(d_{i1} = 1 \mid Y_i)$. This expected value or probability can be evaluated by using Bayes' Rule which, when applied to $E(d_{i1}|Y_i)$ yields

$$P(d_{i1} = 1 \mid Y_i) = \frac{P(d_{i1} = 1)P(Y_i \mid d_{i1} = 1)}{\sum_{j=1}^{2} P(d_{ij} = 1)P(Y_i \mid d_{ij} = 1)} = \frac{\theta f_{i1}}{\theta f_{i1} + (1-\theta) f_{i2}} = w_{i1}. \tag{10}$$

Evaluation of (10) provides estimates of the expected values or probabilities or weights, w_{i1} and $1 - w_{i1}$. Once these weights have been calculated, they can be substituted into the log of the complete-data likelihood which is then maximized in the M step of the EM algorithm with respect to the unknown parameters in the model. To obtain the expressions for evaluation in the EM algorithm, let W_1 and W_2 be given by

$$\begin{aligned} W_1 &= \mathrm{diag}[w_{11}, w_{21}, \ldots, w_{n1}] \\ W_2 &= \mathrm{diag}[w_{12}, w_{22}, \ldots, w_{n2}]. \end{aligned} \tag{11}$$

Clearly, $w_{i1} = 1 - w_{i2}$, for all I, so $W_1 = I_n - W_2$. Differentiating the expected loglikelihood function and solving yields

$$\begin{aligned} \hat{\beta}_1 &= (X'W_1X)^{-1}X'W_1Y \\ \hat{\beta}_2 &= (X'W_2X)^{-1}X'W_2Y \\ \hat{\sigma}_1^2 &= \frac{1}{\sum_{i=1}^{n} w_{i1}} \sum_{i=1}^{n} w_{i1}(Y_i - X_i\hat{\beta}_1)^2 \\ \hat{\sigma}_2^2 &= \frac{1}{\sum_{i=1}^{n}(1 - w_{i1})} \sum_{i=1}^{n} (1 - w_{i1})(Y_i - X_i\hat{\beta}_1)^2 \\ \hat{\theta} &= \sum_{i=1}^{n} w_{i1}. \end{aligned} \tag{12}$$

These solutions are the familiar WLS expressions for the regression parameters in the case of maximum likelihood estimation via the EM algorithm. Given starting values, this algorithm can be used to generate a convergent sequence of parameter estimates.

A Mixture of Stochastic Frontiers. Maximum likelihood estimation via the EM algorithm for a mixture of stochastic frontiers requires that the two EM algorithms for estimating stochastic frontiers and normal mixtures be combined into a single EM algorithm. In the mixture of stochastic frontiers model there are two latent variables; one is the sample separation indicator and the other is the frontier cost. The new EM algorithm must provide expectations for these two variables. This is ultimately achieved by the insertion of a weighting matrix, like that in the mixture-of-normals algorithm, into the stochastic frontier expressions.

For a mixture of stochastic frontiers the density function of the complete-data likelihood is given by

$$f(d_{i1}, Y_i, Z_i) = \left\{\frac{\theta}{\pi\sigma_{v1}\sigma_{w1}} \exp\left\{-\frac{1}{2}\left(\frac{Z_i - X_i\beta_1}{\sigma_{w1}}\right)^2 - \frac{1}{2}\left(\frac{Y_i - Z_i}{\sigma_{v1}}\right)^2\right\}\right\}^{d_{i1}} \cdot \left\{\frac{1-\theta}{\pi\sigma_{v2}\sigma_{w2}} \exp\left\{-\frac{1}{2}\left(\frac{Z_i - X_i\beta_2}{\sigma_{w2}}\right)^2 - \frac{1}{2}\left(\frac{Y_i - Z_i}{\sigma_{v2}}\right)^2\right\}\right\}^{1-d_{i1}}. \tag{13}$$

The logarithm of the complete-data likelihood function is

$$\ln L = \sum_{i=1}^{n}\{d_{i1}(\ln\theta + \ln g_{i1}) + (1-d_{i1})(\ln(1-\theta) + \ln g_{i2})\}, \tag{14}$$

where the gs represent the component stochastic frontier density functions. The EM algorithm replaces both d_{i1} and Z_i by their respective expectations, given Y_i. These calculations are greatly simplified by assuming $E[dZ \mid Y] = E[d|Y]E[Z|Y]$ and $E[dZ^2 \mid Y] = E[d|Y]E[Z^2|Y]$.[6] Then, analogous to the previous result on mixtures, $E(d_{i1}|Y_i) = P(d_{i1} = 1 \mid Y_i)$ is given by

$$P(d_{i1} = 1 \mid Y_i) = \frac{P(d_{i1} = 1)P(Y_i \mid d_{i1} = 1)}{\sum_{j=1}^{k} P(d_{ij} = 1)P(Y_i \mid d_{ij} = 1)} = \frac{\theta g_{i1}}{\theta g_{i1} + (1-\theta)g_{i2}} = w_{i1}. \tag{15}$$

Also, $E(Z_i|Y_i)$ is needed to complete the E step. These expectations are obtained in the same manner as described in equation (6).

If derivatives are taken of the expected loglikelihood equation above, and solved, one obtains

$$\begin{aligned}\hat{\beta}_1 &= (X'W_1X)X'W_1Y \\ \hat{\beta}_2 &= (X'W_2X)X'W_2Y\end{aligned} \tag{16}$$

[6] Following Hartley (1978), the mixture of stochastic frontier regressions can be regarded as a three-equation system consisting of two stochastic frontier equations and a choice equation, in our case containing no explanatory variables. The calculation of these conditional expectations is greatly simplified by assuming that the disturbance term in the choice equation is independent of each of the two-sided error terms in the stochastic frontier equations (these error terms are usually associated with factors outside firm control or "luck"). It is reasonable to assume that a higher than expected likelihood of being in a particular regime is unrelated to deviations from frontier cost or "luck." I am grateful to an anonymous referee for this insight.

$$\sigma_{v1}^2 = \frac{1}{\sum_{i=1}^n w_{i1}} \sum_{i=1}^n w_{i1}(Y_i^2 - 2E(Z_i|Y_i)Y_i + E(Z_i^2|Y_i))$$

$$\sigma_{v2}^2 = \frac{1}{\sum_{i=1}^n (1 - w_{i1})} \sum_{i=1}^n (1 - w_{i1})(Y_i^2 - 2E(Z_i|Y_i)Y_i + E(Z_i^2|Y_i))$$

$$\sigma_{w1}^2 = \frac{1}{\sum_{i=1}^n w_{i1}} \sum_{i=1}^n w_{i1}(E(Z_i^2|Y_i) - 2E(Z_i|Y_i)X_i\beta_1 + (X_i\beta_1)^2)$$

$$\sigma_{w2}^2 = \frac{1}{\sum_{i=1}^n (1 - w_{i1})} \sum_{i=1}^n (1 - w_{i1})(E(Z_i^2|Y_i) - 2E(Z_i|Y_i)X_i\beta_2 + (X_i\beta_2)^2)$$

$$\hat{\theta} = \sum_{i=1}^n w_{i1}.$$

These equations form the basis of the EM algorithm which can be used to estimate parameters in a mixture of normal/half normal stochastic frontier regression models. Given starting values, the algorithm can be iterated until convergence is obtained. After convergence, variance estimates are once again obtained by using the converged parameter values in a single iteration of the algorithm of Berndt, Hall, Hall, and Hausman (1974). In the BHHH method the covariance matrix is estimated by taking the inverse of the sum of the outer products of the gradient vectors, evaluated at the converged parameter values. The virtue of the BHHH method is that only first derivatives of the loglikelihood function are required.

3. Data and model

In a statistical cost function, the dependent variable is total cost and the independent variables are output quantities and input prices. Time is also included as an independent variable, using a suggestion of Berndt (1991). The inclusion of time allows for technical innovations, such as improvements in computing power and software, that presumably can improve the efficiency of any technology.

The selection and specification of regression variables generally follows LeCompte and Smith (1990) and Beard, Caudill and Gropper (1991), and is consistent with the "intermediation" model of the financial firm in which savings and loans take deposits and other inputs and produce loans and related outputs. The three outputs specified are the volume of mortgage loans (MORTGAGE), other loans (OTLOAN), and investments (INVEST) in thousands of real 1986 dollars. Total costs in real 1986 dollars (TOTCOST) are taken as the sum of interest, labor, and capital costs resulting in three input prices. Average real wages (WAGE) by year are obtained from Bureau of Labor Statistics state wage surveys for financial workers. Interest costs per annum on deposits and other loanable fund sources (INTRATE) and physical capital costs (PCAP), inclusive of rent, depreciation, and maintenance, are obtained for each institution for each sample year from the Thrift Financial

Table 1. Regression variables: Descriptive statistics

Variables	Definition	Sample Mean	Sample Standard Deviation
TOTCOST	The sum of interest, labor, and capital costs in thousands of 1986 dollars	28,212	131,954
MORTGAGE	Total volume of mortgage loans in a year in thousands of 1986 dollars	168,160	815,712
OTLOAN	Total volume of loans other than mortgages in a year in thousands of 1986 dollars	15,658	77,197
INVEST	Total volume of investments in a year in thousands of 1986 dollars	60,452	308,301
WAGE	Average annual earnings workers in the financial sector by state in 1986 dollars	24,705	4,401
CAPITAL	Annual real cost per thousand 1986 dollars of physical capital including rent, depreciation, and maintenance	.51	.78
INTRATE	Annual real cost per thousand 1986 dollars borrowed	.073	.008

Source: Thrift Financial Report Tapes, Office of Thrift Supervision, various years, and Bureau of Labor Statistics Wage Surveys, various years (1987–1989).

Report Tapes of the Office of Thrift Supervision. Our model uses information on three outputs, three input prices, and time.

Only solvent FSLIC insured savings and loan associations are included in the analysis. The raw data were screened for a variety of problems, including some institutions that reported negative expenditures, several where total assets did not equal total liabilities plus capital accounts, and others with missing values for any of our regression variables. Also omitted were institutions in Alaska, Hawaii, Guam, and Puerto Rico. Further, no sample S&Ls switched to bank charters during the sample period.

The final data set contained 1,113 institutions in each year, so that the panel contained 3,339 observations. Forty-four of the forty-eight contiguous United States were represented; there were no institutions from Idaho, Mississippi, North Dakota or Wyoming that met all our criteria. The institutions analyzed are widely distributed geographically and exhibited substantial variations in size and output mix. Descriptive statistics are given in Table 1.

The popular translog (transcendental logarithmic) form is used for all the cost estimations. The resulting equation form is:

$$\begin{aligned}\ln C = \alpha_0 &+ \sum \alpha_j \ln q_j + \sum \beta_k \ln p_k + (1/2) \sum\sum \alpha_{ij} \ln q_i \ln qsubj \\ &+ (1/2) \sum\sum \beta_{lk} \ln p_l \ln p_k + \sum\sum \delta_{jk} \ln q_j \ln p_k \\ &+ \psi_1 \ln t + \psi_2 (\ln t)^2 + \sum \psi_{tk} \ln t \ln p_k + \sum \psi_{tj} \ln t \ln q_j,\end{aligned} \tag{17}$$

where C is total cost, q's are output levels, p's are input prices, t is calendar time, and the α's, β's, δ's, and Ψ's are parameters to be estimated. Homogeneity in input prices requires $\sum \beta_k = 1$, $\sum \beta_{lk} = \sum \beta_{kl} = 0$ over 1 and k, $\sum \delta_{jk} = 0$ for any q_j and $\sum \delta_{tk} = 0$. These restrictions are imposed in the estimation. In particular, all prices and quantities are normalized (divided) by the price of capital (PCAP). Also, data are mean-scaled (divided by their means) in order to facilitate calculation of scale economies.

4. Empirical results

The results from estimating the regression models discussed previously are given in Table 2. The second column of Table 2 presents the results of estimating the usual normal-half normal stochastic frontier regression model. The model clearly fits the data well, as is usually the case with cost functions estimated using data from financial institutions. When estimated by OLS, the translog cost function used in all estimations produced an R^2 of 0.986. For the stochastic frontier results in column 2, twenty-five of the twenty-eight regression coefficients are significant at the $\alpha = .10$ level, or better. The coefficients on the output variables are all positive, as expected, and highly significant. The fact that the sum of the output coefficients is less than one suggests that, at the mean output vector, there are scale economies. A t-ratio of 4.06 indicates that the null hypothesis of constant returns to scale can be rejected in favor of the alternative of increasing returns to scale at any of the usual levels of significance. The coefficients of the input price variables are positive, as expected, and the coefficient of the interest rate is highly significant. The negative and significant coefficient of time indicates that costs have generally been falling over the time period.

In the estimation of a stochastic frontier model the variance parameters are also important. In the stochastic frontier model the estimates of σ_v and σ_w are 0.241 and 0.104, respectively. These results show that deviations from the frontier due to the inefficiency error are much larger (2.3 times as great) than deviations due to factors outside firm control.

The results of estimating a mixture of two normal regressions are given in columns 3 and 4 of the table. The mixing parameter is equal to .16 indicating that sixteen percent of the observations are associated with regime 1 and eighty-four percent of the observations are associated with regime 2. Wolfe (1971) developed an adjustment to the usual likelihood ratio statistic for testing for the presence of a mixture. The value of the test statistic in this case is 1454.68, which far exceeds the critical values of a chi-squared distribution with four degrees of freedom at any of the usual levels of significance. Thus, there is strong statistical evidence for the presence of a mixture. That is, the data are better explained by two cost functions than one cost function.

For the regime 1 results given in column 3, only eleven of the twenty-eight regression coefficients are significant at the $\alpha = .10$ level, or better. Still, the coefficients on the output variables are all positive, as expected, and highly significant. The sum of the output coefficients is also still less than one, again suggesting that, at the mean output vector, there are scale economies. A t-ratio of 2.52 indicates that the null hypothesis of constant returns to scale can be rejected in favor of the alternative of increasing returns to scale at the $\alpha = .05$ level of significance. The coefficients of the input price variables are again positive, but only the coefficient of the interest rate is statistically significant. The negative and insignificant coefficient of time indicates that the null hypothesis of no change in costs over the period cannot be rejected.

Table 2. Estimation results

	Stochastic Frontier	Mixture of Normal Regressions		Mixture of Stochastic Frontier Regressions	
VARIABLE	NHN FRONTIER	REGIME 1	REGIME 2	REGIME 1	REGIME 2
INTERCEPT	−0.136 (16.17)[a]	0.078 (2.01)	−0.016 (2.47)	−0.332 (9.42)	−0.090 (7.98)
MORTGAGE	0.591 (118.71)	0.573 (22.76)	0.557 (128.56)	0.590 (31.95)	0.559 (126.95)
OTLOAN	0.097 (24.89)	0.119 (6.54)	0.081 (25.59)	0.184 (8.76)	0.076 (24.22)
INVEST	0.299 (71.78)	0.252 (11.45)	0.363 (95.18)	0.164 (7.10)	0.368 (94.32)
TIME	−0.052 (3.79)	−0.063 (0.75)	−0.039 (3.37)	−0.039 (0.57)	−0.039 (3.42)
WAGE	0.045 (1.71)	0.241 (1.47)	0.024 (1.05)	0.182 (1.24)	0.047 (2.15)
INTRATE	0.956 (35.92)	0.714 (4.17)	0.991 (41.48)	0.834 (5.55)	0.964 (42.25)
SQTINV	0.077 (88.25)	0.052 (9.78)	0.200 (75.78)	0.047 (8.88)	0.198 (70.86)
SQTLOAN	0.016 (14.85)	0.019 (3.88)	0.019 (17.39)	0.070 (7.67)	0.013 (15.33)
SQMORT	0.159 (72.59)	0.140 (12.97)	0.204 (64.38)	0.162 (18.60)	0.200 (63.68)
LOANINV	0.012 (9.65)	0.011 (1.78)	−0.004 (2.85)	−0.004 (0.54)	0.000 (0.01)
LOANMORT	−0.024 (17.21)	−0.021 (2.83)	−0.016 (10.03)	−0.045 (5.92)	−0.012 (7.62)
INVMORT	−0.115 (72.21)	−0.082 (11.12)	−0.192 (86.15)	−0.080 (9.70)	−0.195 (89.61)
SQTIME	−0.070 (1.52)	−0.093 (0.42)	−0.037 (0.95)	0.105 (0.62)	−0.051 (1.27)
SQWAGE	−0.259 (3.14)	0.037 (0.111)	−0.444 (5.47)	0.243 (0.83)	−0.481 (6.16)
SQINT	−0.198 (2.05)	0.174 (0.39)	−0.511 (5.73)	0.613 (1.82)	−0.538 (6.15)
WGETIM	−0.041 (1.20)	0.004 (0.02)	−0.076 (2.46)	0.021 (0.15)	−0.074 (2.38)
WGEINT	0.218 (2.51)	−0.097 (0.26)	0.467 (5.63)	−0.427 (1.42)	0.500 (6.21)

Table 2 (continued)

	Stochastic Frontier	Mixture of Normal Regressions		Mixture of Stochastic Frontier Regressions	
VARIABLE	NHN FRONTIER	REGIME 1	REGIME 2	REGIME 1	REGIME 2
TIMINT	0.028 (0.82)	−0.013 (0.08)	0.073 (2.33)	0.017 (0.12)	0.068 (2.14)
WGEINV	−0.025 (2.38)	−0.002 (0.03)	0.017 (1.44)	0.000 (0.00)	0.019 (1.53)
WGELOAN	−0.063 (6.19)	−0.029 (0.63)	−0.064 (6.73)	−0.108 (2.34)	−0.051 (5.38)
WGEMORT	0.066 (5.07)	0.048 (0.71)	0.024 (1.73)	0.047 (0.75)	0.019 (1.36)
TIMINV	−0.021 (5.74)	0.003 (0.13)	−0.008 (1.66)	0.002 (0.11)	−0.007 (1.50)
TIMLOAN	−0.005 (1.20)	−0.012 (0.49)	−0.005 (1.31)	−0.009 (0.37)	−0.006 (1.67)
TIMMORT	0.024 (3.96)	−0.009 (0.25)	0.015 (2.63)	0.002 (0.08)	0.015 (2.74)
INTINV	0.013 (1.13)	−0.004 (0.07)	−0.010 (0.83)	−0.005 (0.11)	−0.015 (1.20)
INTLOAN	0.062 (5.75)	0.033 (0.65)	0.061 (6.21)	0.097 (2.03)	0.051 (5.23)
INTMORT	−0.047 (3.40)	−0.038 (0.55)	−0.021 (1.50)	−0.011 (0.18)	−0.018 (1.27)
σ		0.233 (57.15)	0.113 (239.14)		
θ		0.160 (10.70)	0.840	0.180 (10.72)	0.820
σ_v	0.241 (46.33)	[0.332][b]	[0.075]	0.337 (23.28)	0.112 (7.13)
σ_w	0.104 (28.11)	[0.119]	[0.103]	0.056 (3.52)	0.094 (9.87)
−LOGL	1104.53	1720.92		1757.96	

[a] Numbers in parentheses are the absolute values of *t*-ratios.
[b] The numbers in brackets are obtained by using the relationship for adjusted OLS estimate.

The regime 2 results, given in column 4 and representing eighty-four percent of the observations, are somewhat better. Twenty-two of the twenty-eight regression coefficients are significant at the $\alpha = .10$ level, or better. The coefficients on the output variables are, again, all positive and highly significant.

The sum of the output coefficients is very close to one, suggesting the presence of constant returns to scale. A t-ratio of 0.38 indicates that the null hypothesis of constant returns to scale can not be rejected at any of the usual levels of significance. The coefficients of the input price variables are again positive, but only the coefficient of the interest rate is statistically significant. The negative and significant coefficient of time again indicates that costs have generally declined over the period.

The converged parameter values from estimating a mixture of two stochastic frontiers are contained in columns 5 and 6 of Table 2.[7] The value of the mixing parameter is .18 indicating that eighteen percent of the observations are associated with regime 1 and eighty-two percent of the observations are associated with regime 2. Applying the Wolfe test for the presence of a mixture yields an approximate chi-square value of 1308.17, which far exceeds the critical values for a chi-squared distribution with four degrees of freedom at the usual levels of significance. Thus, there is strong statistical evidence for the presence of a mixture of stochastic frontiers.

The stochastic frontier regime 1 results, given in column 5, represent only eighteen percent of the observations. Only thirteen of the twenty-eight regression coefficients are significant at the $\alpha = .10$ level, or better. Even so, the coefficients on the output variables are all positive, as expected, and significant. The sum of the output coefficients is again less than one, suggesting the presence of scale economies at the mean output vector. A t-ratio of 3.65 indicates that the null hypothesis of constant returns to scale can be rejected in favor of the alternative of increasing returns to scale at any of the usual levels of significance. The coefficients of the input price variables are, again, positive, but only the coefficient of the interest rate is statistically significant. The negative and insignificant coefficient of time indicates that the null hypothesis of no change in costs over the period cannot be rejected.

The estimates of the variance parameters, σ_v and σ_w, are 0.337 and 0.056, respectively. These results show that deviations from the frontier due to the inefficiency error are much larger than deviations due to factors outside firm control.

The stochastic frontier regime 2 results given in column 6 represent the remaining eighty-two percent of the observations. Twenty-one of the twenty-eight regression coefficients are significant at the $\alpha = .10$ level, or better. The coefficients on the output variables are all positive, and, again, highly significant. The sum of the output coefficients is, again, very near one, suggesting the presence of constant returns to scale at the mean output vector. A t-ratio of 1.00 indicates that the null hypothesis of constant returns to scale can not be

[7] The program used to estimate the models is written in the matrix language IML in SAS. Convergence is examined in two steps. Initially, the convergence is determined by examining the value of the loglikelihood function and the value of the mixing parameter. Then, when these values have ceased to change, the gradient and the adjustment vector are examined. When elements of these vectors are suitably small, convergence is declared. For example, in the estimation of the mixture of stochastic frontier regressions, the values of the loglikelihood function and mixing weight have stopped changing, the norm of the gradient is .23888, the norm of the adjustment vector is .0000921, the maximum of the absolute values of the elements in the gradient vector is .12214, and the maximum of the absolute values of the elements in the adjustment vector is .0000411. These results are typical of the estimations in the paper.

rejected at any of the usual levels of significance. The coefficients of the input price variables are again positive, and both are statistically significant. The negative and significant coefficient of time indicates once more that costs have been declining over the period.

For regime 2, the estimates of the variance parameters, σ_v and σ_w, are 0.112 and 0.094, respectively. These results show that deviations from the frontier due to the inefficiency error are very similar in size to deviations from the frontier due to factors outside firm control.

At this point it is quite clear that the two mixture regimes and the two stochastic frontier regimes correspond rather closely to one another. Directly comparing the estimation results for these models is tedious due to the large number of parameters estimated. However, some useful comparisons can easily be made by either examining variance measures from each model or by comparing the predicted costs from each.

A simple comparison can be made by examining the standard deviations of the components of the variance. As noted earlier, the estimate of σ_v for regime 1 is 0.337 and the estimate of σ_v for regime 2 is 0.112. These figures suggest that there is about three times as much variability in the inefficiency associated with regime 1. The estimate of σ_w for regime 1 is 0.056 and the estimate of σ_w for regime 2 is 0.094. These figures suggest that there is nearly twice as much variability in the random error associated with regime 2.

Comparisons can also be facilitated by examining predicted costs. Predicted costs for each of the five models discussed are given in Table 3. Costs are calculated at several different percentages of the mean output vector. The

Table 3. Predicted costs and percent of mean predicted cost

Percent of Mean Output	NHN FRONTIER	MIX1	MIX2	MIXF1	MIXF2
50%	13203.0 (53.5%)[a]	16795.7 (55.0%)	15342.9 (55.3%)	11359.6 (56.1%)	14196.7 (55.0%)
100%	24627.2 (100.0%)	30516.4 (100.0%)	27755.8 (100.0%)	20236.6 (100.0%)	25795.7 (100.0%)
150%	37520.2 (152.4%)	45619.4 (149.5%)	43119.5 (155.4%)	30346.6 (150.0%)	40052.8 (155.3%)
200%	51874.8 (210.5%)	62129.8 (203.6%)	61467.9 (221.5%)	41693.8 (206.0%)	56994.7 (220.9%)
250%	67662.6 (274.4%)	80023.8 (262.2%)	82889.8 (298.6%)	54272.0 (268.2%)	76690.6 (297.3%)
300%	84858.0 (344.6%)	98450.2 (322.6%)	107542.0 (387.5%)	68075.6 (336.4%)	99220.4 (384.6%)
400%	123395.0 (501.1%)	140320.2 (459.8%)	166741.4 (600.7%)	99350.1 (490.9%)	153127.0 (593.6%)
500%	167373.7 (679.6%)	187197.8 (613.4%)	239993.5 (864.7%)	135520.8 (669.7%)	219432.1 (850.7%)

[a] Figures in parentheses are percentages of mean cost.

table also shows that the predicted costs from regimes 1 and 2 in the mixture of normal regressions model are very similar. At fifty percent of the mean output vector, regime 2 costs are lower than regime 1. As output increases, the distance between these costs initially increases, but ultimately declines. At two hundred fifty percent of the mean output vector, regime 2 costs actually exceed regime 1 costs. The table also shows that predicted costs from the stochastic frontier model are below the predicted costs from every model except the stochastic frontier regime 1 model.

The predictions from the mixture of stochastic frontiers are very interesting. The cost predictions from regime 1 are the lowest in the table, even lower than those from the stochastic frontier. Predicted costs from stochastic frontier regimes 1 and 2 lie below corresponding mixture regimes 1 and 2. Even the high cost stochastic frontier regime 2 lies below both mixture regimes. These cost functions are ranked, generally, from lowest cost to highest cost, as; first, stochastic frontier regime 1, second, stochastic frontier, third, stochastic frontier regime 2, fourth, mixture regime 1, and fifth, mixture regime 2.

These cost predictions, along with the earlier inefficiency results, suggest that there seem to be a small number of savings and loans characterized by very low costs, with some firms being very inefficient with respect to that technology. Still, the relatively inefficient firms appear to be operating with lower costs than the bulk of firms using the other technology. Most of the firms are relatively more efficient while using a higher cost technology. A possible explanation can be found in technology switching behavior. Some firms, having made the switch to a more efficient technology some time in the past, are operating very efficiently. Firms more recently switching to the more efficient technology may have high costs for a time. Separating firms into regimes may help provide an explanation.

The first step in comparing the two regimes from the mixture of stochastic frontiers is to separate the observations by regime. This is possible by using the weighting matrix calculated as part of the EM algorithm. An observation is assigned to regime 1 if the weight for that regime exceeds one half, otherwise the observation is assigned to regime 2. This assignment scheme results in 244 observations assigned to regime 1 and 3095 observations assigned to regime 2.

The transformed and untransformed means for the total sample and for each of the two regimes are given in Table 4. The untransformed means are provided because they are easy to interpret. The transformed means are provided because they represent the data used in the actual estimations. The means calculated for the regimes reveal some distinct differences between the two regimes. The first large difference revealed in the table is between the total costs for regimes 1 and 2. The average total cost for savings and loans in regime 1 is \$16,883,000 but for savings and loans in regime 2 the average is \$29,106,000 or 72 percent higher. The average volume of mortgages is almost 82 percent higher in regime 2 than in regime 1. The same is true with investments. The average volume of investments is almost 57 percent higher in regime 2. These facts lead to the conclusion that the institutions associated with regime 2 are considerably larger than those associated with regime 1.

Taken together, these results certainly imply that one cost function does not adequately describe the savings and loan data. The empirical evidence

Table 4. Raw and transformed means [Transformed means in brackets]

VARIABLE	TOTAL	REGIME 1	REGIME 2
TOTCOST	28212 [−0.955]	16883 [−1.334]	29106.2 [−0.925]
MORT	168160 [−1.283]	95643.5 [−1.786]	173877 [−1.243]
OTLOAN	15658 [−2.012]	16252.7 [−2.125]	15611.2 [−2.003]
INVEST	60452 [−1.659]	39581.6 [−2.569]	62098.3 [−1.587]
WAGE	24705 [0.333]	25069.2 [0.181]	24676 [0.345]
INTRATE	0.073 [0.342]	0.072 [0.148]	0.074 [0.357]

suggests that smaller institutions are operating along a different cost function than larger ones. Perhaps this difference is due to differences in the underlying technology, differences in the structures of the markets in which these institutions operate, differences in product mix, or some combination of these factors. What is clear is that a single frontier cost function does not adequately describe the data. A solution to this problem is to estimate a mixture of stochastic frontiers using the EM algorithm developed in this paper.

Conclusions

This paper extends previous work in the area of cost estimation and efficiency by estimating a mixture of two stochastic frontiers in the presence of *no* sample separation information. An EM algorithm, based primarily on the work of Huang (1984) and Hartley (1978), is developed for the estimation of a mixture of stochastic frontiers.[8] This method can be used when one suspects there is more than one stochastic frontier generating the data, but one has no information on which observations are associated with which regime. The algorithm developed in this study makes possible the estimation of stochastic frontiers with no sample separation information.

The EM algorithm developed is applied to the estimation of a mixture of two stochastic cost frontiers using data on solvent savings and loans for the period 1986–1988. The results indicate that the cost data are characterized by a mixture of stochastic frontiers.

[8] A more detailed version of the paper is available from the author upon request.

References

Aigner DJ, Amemiya T, Poirier DJ (1976) On the estimation of production frontiers: Maximum likelihood estimation of the parameters of a discontinuous density function. International Economic Review 17:377–396

Aigner D, Lovell CAK, Schmidt P (1977) Formulation and estimation of stochastic frontier production function models. Journal of Econometrics 6:21–37

Battese G, Corra G (1977) Estimation of a production frontier model: With application to the pastoral zone of eastern Australia. Australian Journal of Agricultural Economics 21:169–179

Beard TR, Caudill SB, Gropper DM (1991) Finite mixture estimation of multiproduct cost functions. Review of Economics and Statistics 73:654–664

Beard TR, Caudill SB, Gropper DM (1997) The diffusion of production processes in the U.S. banking industry: A finite mixture approach. Journal of Banking and Finance 21:721–740

Beckers D, Hammond C (1987) A tractable likelihood function for the normal-Gamma stochastic frontier model. Economics Letters 24:33–38

Berger AN (1993) 'Distribution-free' estimates of efficiency in the U.S. banking industry and tests of the standard distributional assumptions. Journal of Productivity Analysis 4:261–292

Berger AN, Mester (1997) Inside the black box: What explains differences in the efficiencies of financial institutions? Journal of Banking and Finance 21:895–947

Berndt E (1991) The practice of econometrics. Reading, Mass.: Addison-Wesley

Berndt E, Hall B, Hall R, Hausman J (1974) Estimation and inference in nonlinear statistical models. Annals of Economic and Social Measurement 3:653–666

Caudill SB (1993) Estimating the cost of partial-coverage rent controls: A stochastic frontier approach. The Review of Economics and Statistics 75:727–31

Dempster AP, Laird N, Rubin D (1977) Maximum likelihood estimation from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B, 39:1–38

Douglas SM, Conway KS, Ferrier GD (1995) A switching frontier model for imperfect sample separation information: With an application to constrained labor supply. International Economic Review 36:503–526

Everitt BS, Hand DJ (1981) Finite mixture distributions. New York: Chapman & Hall

Fair RC, Jaffee DM (1972) Methods of estimation for markets in disequilibrium. Econometrica 40:497–514

Goffe WL, Ferrier GD, Rogers J (1994) Global optimization of statistical functions with simulated annealing. Journal of Econometrics 60:65–99

Goldfeld SM, Quandt RE (1976) Techniques for estimating switching regressions. In Studies in Nonlinear Estimation

Greene W (1990) A Gamma-distributed stochastic frontier model. Journal of Econometrics 13:101–115

Hartley M (1978) Comment (on "Estimating Mixtures of Normal Distributions and Switching Regressions," by Quandt and Ramsey). Journal of the American Statistical Association 73:738–741

Huang CJ (1984) Estimation of stochastic frontier production function and technical inefficiency via the EM algorithm. Southern Economic Journal 50:847–856

Jondrow J, Lovell CAK, Materov IS, Schmidt P (1982) On the estimation of technical in inefficiency in the stochastic frontier production function model. Journal of Econometrics 23:269–274

LeCompte RLB, Smith SD (1990) Changes in the cost of intermediation: The case of savings and loans. Journal of Finance 45:1337–1346

Lee LF (1983) On maximum likelihood estimation of stochastic frontier production models. Journal of Econometrics 23:269–274

Lee LF, Porter RH (1984) Switching regression model with imperfect sample separation information: With an application on cartel stability. Econometrica 52:391–418

Maddala GS (1983a) Limited-dependent and qualitative variables in econometrics. Cambridge: Cambridge University Press

Maddala GS (1983b) Methods of estimation for models of markets with bounded price variation. International Economic Review 24:361–378

Maddala GS, Nelson FD (1974) Maximum likelihood methods for models of markets in disequilibrium. Econometrica 42:1013–30

McLachlan G, Basford K (1988) Mixture models: Inference and applications to clustering. New York: Marcel Dekker

McLachlan G, Krishnan T (1997) The EM algorithm and extensions. New York: John Wiley

Meeusen W, Van den Broeck J (1977) Efficiency estimation from Cobb-Douglas productions with composed error. International Economic Review 18:435–444

Mester LJ (1993) Efficiency in the savings and loan industry. Journal of Banking and Finance 17:267–286

Polachek SW, Yoon BJ (1987) A two-tiered earnings frontier estimation of employer and employee information in the labor market. The Review of Economics and Statistics 69:296–302

Quandt R (1972) A new approach to estimating switching regressions. Journal of the American Statistical Association 67:306–310

Quandt R (1988) The econometrics of disequilibrium. New York: Basil Blackwell

Quandt R, Ramsey J (1978) Estimating mixtures of normal distributions and switching regressions. Journal of the American statistical Association 73:735–747

Stewart M (1983) On least squares estimation when the dependent variable is grouped. Review of Economic Studies 50:141–149

Stevenson R (1980) Likelihood functions for generalized stochastic frontier estimation. Journal of Econometrics

Titterington M, Smith A, Makov V (1985) Statistical analysis of finite mixture distributions. New York: John Wiley and Sons

Wolfe JH (1971) A Monte Carlo study of the sampling distribution of the likelihood ratio for mixtures of normal distributions. Technical Bulletin STB 72. San Diego: U.S. Naval Personnel and Training Research Laboratory.