ECONOMIC DISCUSSION PAPERS

Efficiency Series Paper 11/2002

Measuring Efficiency using a Stochastic Frontier Latent Class Model

Luis Orea y Subal Kumbhakar

Departamento de Economía

Universidad de Oviedo

UNIVERSIDAD DE OVIEDO

DEPARTAMENTO DE ECONOMÍA

PERMANENT SEMINAR ON EFFICIENCY AND PRODUCTIVITY

MEASURING EFFICIENCY USING A STOCHASTIC FRONTIER LATENT CLASS MODEL

Luis Orea ♠ **and Subal C. Kumbhakar** ♣

Efficiency Series Paper 2002/11

Abstract
Efficiency estimation in stochastic frontier models typically assumes that the underlying production technology is the same for all firms. There might, however, be unobserved differences in technologies and input/output qualities that can be inappropriately labeled as inefficiency if such differences are not taken into account. We address this issue by developing a Stochastic Frontier Latent Class Model in a "panel data" framework. This model exploits the information contained in the data more efficiently compared to the traditional cluster analysis. An application of the proposed model is presented using Spanish banking data.

Keywords: stochastic cost frontier, latent class model, panel data, bank.

♠ Department of Economics, University of Oviedo, Spain.
E-mails: kkar@binghamton.edu; lorea@correo.uniovi.es

♣ Department of Economics, State University of New York, Binghamton, USA.

1. Introduction

Stochastic production (or cost) frontier functions are now increasingly used to measure efficiency of individual producers. Estimation of these functions rests on the assumption that the underlying production technology is common to all producers. However, firms in an industry may use different technologies. In such a case estimating a commom frontier function encompassing the entire sample observations may not appropriate in the sense that the estimated technology is not likely to represent the 'true' technolgy. That is the estimate of the underlying technology may be baised. In other words, if the unobserved technological differences are not taken into account in estimation, effect of these omitted unobservables might be inappropriately labeled as inefficiency.

To reduce the likelihood of these types of misspecification, researchers often estimate frontier functions by classifying the sample observations into certain categories using exogenous sample separation information. For instance, Mester (1993) and Grifell and Lovell (1997) grouped banks into private and savings banks. Kolari and Zardkoohi (1995) estimated separate costs functions for banks grouped in terms of their output mix. Mester (1997) grouped sample banks in terms of their location. Polachek and Yoon (1987) allowed for different regimes in estimating of the earning frontier functions of employers and employee.

In the above studies, estimation of the technology using a sample of firms is carried out in two stages. First, the sample observations are classified into several groups. This classification is based on either some a priori sample separation information (e.g., ownership of firms (private, public and foreign), location of firms, etc.) or applying cluster analysis to variables such as output and input ratios. In the second stage, separate analyses are carried out for each class/sub-sample. This procedure is not *efficient* in the sense that information contained in one class is not used to estimate the technology (production or cost frontier) of firms that belong to other classes. However, in most of the empirical applications this inter-class information may be quite important because firms belonging to different classes often come from the same industry/sector. Although their technologies may be different, they share some common features.

To exploit the information contained in the data more efficiently, we advocate using a Stochastic Frontier Latent Class Model (hereafter SFLCM) that combines the stochastic frontier approach and a latent class structure.[1] In this model both firm's technology and the probability of particular group membership are estimated *simultaneously*. The frontier that is estimated for one particular class might (with a nonzero probability) be the reference technology for any observation, whether it belongs to that class or to some other classes. This implies that all the observations in the sample should be used to estimate the underlying technology for each class.[2] The proposed methodology also classifies the sample into several groups even when sample-separating information is not available (which is required in a traditional cluster analysis). In this case, the latent class structure uses the goodness of fit of each estimated frontier as additional information to identify groups of firms.

Recently only a few studies combined the stochastic frontier approach with the latent class structure in order to estimate a mixture of frontiers functions.[3] In particular, Caudill (2002) introduces an expectation-maximization (EM) algorithm to estimate a mixture of two stochastic cost frontiers in presence of no sample separation information.[4] Greene (2001) proposes a maximum likelihood SFLCM using sample separation information and allowing for more than two classes, although he does not provide any application.

The main feature of the models proposed by both Caudill and Greene is that they assume independence of the efficiency term over time. That is, their models are developed in a "cross-sectional" framework where a firm observed in two periods is treated as two separate firms. This assumption doesn't allow one to test whether the efficiency is time-invariant or not. In any case, time-invariant inefficieny is not

---

[1] See Greene (2002) for a survey of latent class models.

[2] In the standard procedure, we are *implicitly* restricting the cross-class probabilities to be zero and the own probabilities to be equal one. This precludes using observations from other classes to estimate a particular class frontier.

[3] This combination has also been employed by Tsionas (2000) to allow for heterogeneity in the distribution of the inefficiency term, exclusively.

[4] See, in addition, Beard, Caudill and Gropper (1991, 1997) for applications using a non-frontier approach.

particularly appealing in a productivity growth study.[5] We avoid this problem by developing a SFLCM in a "panel data" framework, which allows cost (or technical) efficiency to vary over time in a parametric form. An application of the proposed model is presented using Spanish banking data.[6]

The rest of the paper is organized as follows. Section 2 describes the empirical model. Section 3 describes the data. Section 4 reports the empirical results. Section 5 contains a summary and some concluding remarks.

2. Panel data specification of a Stochastic Frontier Latent Class Model

To determine efficiency the technology of banks belonging to each class must be modeled. Here we assume that the techology is represented by the dual cost function. In particular, we assume that the cost fuction for class *j* is of the *translog* form, viz.,

$$\ln C_{it} = \ln C(y_{it}, w_{it}, t, \boldsymbol{b}_j) + u_{it}\,|_j + v_{it}\,|_j \tag{1}$$

where subscripts i = 1,…,N; t =1,…,T; and j =1,…,J stand, for firm, time and class, respectively; C_{it} is actual total cost; y_{it} and w_{it} are, respectively, vectors of outputs and input prices; and β_j is the vector of parameters to be estimated for class *j*. For each class, the stochastic nature of the frontier is modeled by adding a two-sided random error term $v_{it}|_j$, which is assumed to be independent of a non-negative cost inefficiency disturbance $u_{it}|_j$.

Additional restrictions should be imposed in order to estimate (1) by the maximum likelihood method. In particular, the noise term for class *j* is assumed to follow a normal distribution with mean zero and constant variance, s^2_{vj}. The inefficiency term $u_{it}|_j$ is

[5] These models do not estimate the inefficiency term consistently since its variance does not vanish as the sample size increases. A detailed discussion of this issue can be found in Schmidt and Sickles (1984) and Greene (1993a).

[6] The Spanish banking industry has many types of banks, although the savings and private banks are the majority. We distinguish between the savings banks and private banks because they have been regulated in different ways and have been traditionally specialized in different services. A small number of the private banks are universal banks while a large number of them are regional banks that employ, just like savings banks, a high proportion of deposits to fund loans. There are also some non-commercial national and foreign banks that specialize in interbanking activities.

modeled as the product of a time-invariant firm effect, $u_i|_j$, and a parametric function of time $\lambda_t(\eta_j)$, where η_j is a vector of parameters to be estimated. The term $u_i|_j$ is supposed to come from a non-negative truncated normal distribution with zero mean and variance s_{uj}^2.

Several forms for the function λ_t have been proposed in the literature. A common feature of them is that they are *exclusively* functions of time. Following the literature,[7] we adopt an exponential form for λ_t, but allow other variables that might explain differences over time or among firms (e.g., public, private, etc.) to be included in λ_t. In particular, the specification of our inefficiency term $u_{it}|_j$ can be written, in general terms, as:

$$u_{it}\,|_j = \boldsymbol{l}_t(\boldsymbol{h}_j)\cdot u_i\,|_j = \exp\left(z_{it}\cdot \boldsymbol{h}_j{}'\right)\cdot u_i\,|_j \quad , \quad u_i\,|_j \geq 0 \tag{2}$$

where $h_j = (h_{1j},\ldots,h_{Hj})$ are parameters and $z_{it} = (z_{1it},\ldots,z_{Hit})$ is a matrix of H variables that might affect inefficiency. This specification yields several other parametric functions proposed in the literature as special cases. If h_j is a scalar and $z_{it} = (T\text{-}t)$ we get the specification proposed by Battese and Coelli (1992). If h is a 1×2 vector and $z_{it}=(t\ t^2)$ we get the specification proposed by Kumbhakar (1990). Finally, if h_j is a $1\times T$ vector and z_{it} is a set of T time-dummy variables, we get the specification proposed by Lee and Schmidt (1993).

With these distributional assumptions, the log likelihood function for firm i assuming that it belongs to class j can be written as (see Battese and Coelli, 1992):

$$\begin{aligned}\ln LF_{ij}(\boldsymbol{q}_j) = \; & \ln\left[1-\Phi(-z_i^*)\right]+(z_i^*)^2-\frac{1}{2}\left[\ln 2\boldsymbol{p}+\ln \boldsymbol{s}_j^2\right]\cdot T_i-\frac{1}{2}\ln(1-\boldsymbol{g}_j)\cdot(T_i-1)\\ & -\frac{1}{2}\cdot\ln\left[1+\boldsymbol{g}_j\cdot\left(\textstyle\sum_{i=1}^{T_i}\boldsymbol{l}_t(\boldsymbol{h}_j)^2-1\right)\right]-\frac{1}{2}\cdot\sum_{t=1}^{T_i}\left[\boldsymbol{e}_{it}(\boldsymbol{b}_j)^2\big/(1-\boldsymbol{g}_j)\boldsymbol{s}_j^2\right]\end{aligned} \tag{3}$$

where

$$z_i^* = \frac{\boldsymbol{g}_j\cdot\sum_{t=1}^{T_i}\boldsymbol{l}_t(\boldsymbol{h}_j)\cdot\boldsymbol{e}_{it}(\boldsymbol{b}_j)}{\left\{\boldsymbol{g}_j\cdot(1-\boldsymbol{g}_j)\cdot\boldsymbol{s}_j^2\cdot\left[1+\boldsymbol{g}_j\cdot\left(\sum_{t=1}^{T_i}\boldsymbol{l}_t(\boldsymbol{h}_j)^2-1\right)\right]\right\}^{1/2}}$$

---

[7] See the next paragraph.

with $\varepsilon_{it} = \varepsilon_{it}(\beta_j)$ = lnC_{it} - lnC (y_{it}, w_{it}, t, β_j); $s_j^2 = s_{vj}^2 + s_{uj}^2$; $g_j = s_{uj}^2 / s_j^2$; and $q_j = (b_j\ s_j^2\ g_j\ h_j)$ are the parameters associated with the technology of class j, and $\Phi(\cdot)$ is the standard normal distribution function.

Note that the likelihood function in (3) is defined for all the time periods over which firm i is observed, while in Greene it is defined for firm i at each time t. Thus, the full contribution of firm i to the overall likelihood function is obtained in Greene's paper as LF$_i$ = $\prod_t^{T_i} LF_{it}$, where LF_{it} is the likelihood function for firm i at time t. This, however, cannot be done in our model due to the fact that firm observations are not independent over time.

The class determination for each firm is addressed by adopting a latent class structure. In this formulation, the likelihood function for the firm i is obtained as the weighted sum of their j-class likelihood functions, where the weights are the probabilities of class membership. That is,

$$LF_i(\boldsymbol{q},\boldsymbol{d}) = \sum_{j=1}^{J} LF_{ij}(\boldsymbol{q}_j) \cdot P_{ij}(\boldsymbol{d}_j) \quad , \quad 0 \le P_{ij} \le 1 \quad , \quad \Sigma_j P_{ij} = 1 \tag{4}$$

where q = ($q_1,\ldots,q_J$), d = ($d_1,\ldots,d_J$) and the class probabilities are parameterized as a multinomial logit model,

$$P_{ij}(\boldsymbol{d}_j) = \frac{\exp(\boldsymbol{d}_j' q_i)}{\sum_{j=1}^{J} \exp(\boldsymbol{d}_j' q_i)} \quad , \quad j = 1,\ldots,J \quad , \quad \boldsymbol{d}_J = 0 \tag{5}$$

where q_i is a vector of firm-specific variables. The overall likelihood function resulting from (3) to (5) is a continous function of the vectors of parameters θ and δ, and can written in as:

$$\ln LF(\boldsymbol{q},\boldsymbol{d}) = \sum_{i=1}^{N} \ln LF_i(\boldsymbol{q},\boldsymbol{d}) = \sum_{i=1}^{N} \ln\left\{\sum_{j=1}^{J} LF_{ij}(\boldsymbol{q}_j) \cdot P_{ij}(\boldsymbol{d}_j)\right\} \tag{6}$$

Under the mantained assumptions, maximum likelihood techniques will give asymptotically efficient estimates of all the parameters. A *necessary* condition for identifing δ, the parameters of the latent class probabilities, is that the sample must be generated from either different technologies or different noise/inefficiency terms. Otherwise, we cannot classify the observations into several groups since the vector of parameters δ is not identified.[8]

The estimated parameters can be used to compute the conditional posterior class probabilites. Following the steps outlined in Greene (2001) the posterior class probabilities can be obtained from

$$P(j \mid i) = \frac{LF_{ij}(\boldsymbol{q}_j) \cdot P_{ij}(\boldsymbol{d}_j)}{\sum_{j=1}^{J} LF_{ij}(\boldsymbol{q}_j) \cdot P_{ij}(\boldsymbol{d}_j)} \tag{7}$$

This expression shows that the posterior class probabilities depends not only on the estimated δ parameters, but also on the vector θ, i.e., the parameters from the cost frontier. This supports our statement that, in addition to the variables included in the latent class probabilities, a latent class model uses the goodness of fit of frontiers from every class as additional information to identify groups of firms.[9]

In the standard stochastic frontier apporach where the frontier function (the reference technology) is same for every firm, we estimate inefficiency relative to the frontier for all

---

[8] It can be shown that the vector of parameters δ is not identified if firms belonging to different groups use the same technology (i.e., $q_j = q_h$) Consider, for instance, the case of two groups (J = 2). The first derivative of the log likelihood (6) with respect to δ can then be formulated as:

$$\partial \ln LF(q,d)/\partial d = \sum_{i=1}^{N} (LF_{i1}(q_1) - LF_{i2}(q_2))/LF_i \cdot \partial P_{i1}(d)/\partial d$$

this equation shows that irrespective of the value of δ, ∂lnLF/∂δ is always equal to zero if the stochastic cost frontier parameters are the same in both classes.

[9] It is to be noted that although Greene (2001) works with a density function for each firm *i* at time *t*, he proposes estimating the posterior class probability for the *complete* set of observations of firm *i*. That is, as in equation (7), he proposes estimating P(*j*|*i*) instead of P(*j*|*i,t*). This seems to support our strategy of constructing the whole model from the firm's point of view, and not from the density function of each observation *i* at time *t*. This difference does not seems to be important because the expression used here and the one proposed by Greene for estimating P(*j*|*i*) are equivalent, except for $LF_{ij}(q_j)$ which in Greene (2001) is estimated as the product of T_i independent density functions, whereas here it is estimated using equation (3). Since the likelihood functions are different the estimated parameters are likely to be different. Thus although the same formula is used to compute the posterior probabilities – the estimated probabilities are likely to differ.

observation, viz, inefficiency from $E(u_{it} \mid \boldsymbol{e}_i)$ and efficiency from $E[exp(-u_{it})|e_i]$.[10] In the present case, the interpretation of inefficiency is, however, different. The model proposes that a firm may belong to more than one class (with some probability). Thus there is no unique reference technology against which inefficiency is to be computed. There are two ways to solve this problem. First, we examine the posterior probability for each firm and assign it a class based on the highest probability (assuming that there is no tie). Once the class assignment is done inefficiency for that firm is computed using the frontier for the assigned class as its reference technology. Note that this method ignores all other class probabilites although the (posterior) class probabilities are not zero. This scheme of arbitrary weighting and somewhat ad hoc choice of the reference technology can be avoided by using the second method, viz.,

$$\ln EF_{it} = \sum_{j=1}^{J} P(j \mid i) \cdot \ln EF_{it}(j) \tag{8}$$

where P(j|i) is the posterior probability of belonging to class j given firm i defined in (7), and $EF_{it}(j)$ is its efficiency using the technology of class j as the reference technology. Note that here we don't have a single reference technology. It takes into account technologies from every class.

This is the strategy suggested by Greene (2001) to get firm-specific estimates of the parameters of the stochastic frontier model. The efficiency results obtained from using (8) would be different from those based on the most likely frontier and using it as the reference technology. These differences appear due to the fact that the "reference" frontiers are not the same. The size of the difference depends on the relative importance of the posterior probability of the most likely cost frontier, the higher the posterior probability the smaller the differences.

3. Data and sample

This section includes an application of the SFLCM discussed in the previous section using the Spanish private and savings banks. The number of banks decreased steadily

---

[10] ε_i denote the vector of the T_i associated ε_{it} for firm i. For more details, see Kumbhakar (1990) and Battese and Coelli (1992, eq. 3). These authors extend Jondrow et al. (1982) result that allows computation of individual inefficiencies in a panel data framework.

over the last ten years due to merger and adquisitions.[11] Since mergers took place mainly in the early 1990s and in 1992 a change took place in the structure of the public balance sheets that reduced the amount of information reported by banks, we use an unbalanced panel of 169 banks over the period 1992-2000.

Three sets of variables are required to estimate the model introduced in Section 2. These are: the stochastic frontier variables (i.e. C_{it}, y_{it} , t and w_{it}); the z_{it} variables in the parametric function of the inefficiency term; and the q_i variables in the class probabilities.

The variables used in the stochastic cost frontier are defined in the same way for every group of banks. We follow the banking literature and use the intermediation approach proposed by Sealey et al. (1977) to define inputs and outputs. The intermediation approach treats deposits as inputs and loans as outputs. In our application we include four types of outputs, viz., bonds, cash and others assets not covered by the following outputs (y_1); interbanking loans (y_2); loans to firms and households (y_3); and non-interest income (y_4). The last output is not commonly used in the intermediation approach. We include non-interest income in an attempt to capture off-balance-sheet activities such as, brokerage services, management of financial assets or mutual funds for the customers. These activities are becoming increasingly important in Spanish banks.[12]

Total cost includes both interest and operating expenses. The interest expenses explains about 71% of total cost and they came from demand, time and saving deposits, deposits from non-banks, securities sold under agreements to repurchase, and other borrowed money. The operating expenses that represent the remaining 29% of total cost includes labor expenses and other general operating expenses, such as

---

[11] While a merger implies that a new bank is born and the disappearance of two banks, in an acquisition only one disappears and no new bank is born.

[12] Our measure of nontraditional banking activities is not without problems. First, we cannot distinguish between variations due to changes in volumes and variations due to changes in prices. Second, non-interest income is partly generated from traditional activities (such as fees from service charges on deposits or credits) rather than nontraditional activities alone. Since comprehensive information on the degree of off-balance-sheet services is not available, we prefer to describe them in an approximate way. Many recent efficiency studies also include fee or non-interest income as an output (for example, Lang and Welzel (1996), Resti (1997) and Rogers (1998)).

rent and occupancy cost, communication expenses, or travel and relocation expenses. Since comprehensive information about the amount of physical assets and other operational inputs is not available in our database, we do not distinguish between labor and other operational expenses. Accordingly, we include two input prices in our cost functions. These are: loanable funds price, measured by dividing interest expenses by total amount of deposits and other loanable funds (w_1); and operational inputs price, measured by dividing total operating expenses by total number of employees (w_2). The descriptive statistics of these variables can be found in Table 1. All monetary variables were deflated by the GDP deflator index, and are expressed in thousand Euros (using 2000 as the base year).

Regarding the parametric part of the inefficiency term, $l(.)$, we consider three z_{it} variables. The first variable is the time trend (t). With only time the specification of $l(.)$ corresponds to the Battese and Coelli (1992) form. Since $l(.)$ is a function of time with only one parameter, efficiency either increases, decreases or remains constant. The second variable, D_A, is a dummy variable that increases its value by one unit each time the bank extended its activity through acquisition of other bank, and it takes a value of zero if the bank doesn't acquire other finantial institutions from the second year to the last year of the sample. Since an acquisition process involves structural changes (closure of branches, staff relocation, etc.), we expect an increase in inefficiency when an acquisition takes place. The third variable, D_S, is also a dummy variables that takes a value one if the finantial institution is a savings banks, and zero otherwise. The coefficient of this variables allow us to test whether savings banks are as efficient as private banks.

Finally, we consider the firm-average value of five variables, apart from an intercept, as determinants of the latent class probabilities. As customary in cluster analysis, the variables included in the class probabilities are four balance sheet ratios: loans to firms and households (L_{NB}), interbanking loans (L_B), time and saving deposits (D_{NB}), and deposits from banks (D_B). We also include the labor to branch ratio (LBR) to identify a set of non-commercial banks that operate in large population cities with large branches.

In summary the final specification of the cost frontier model (ignoring the *j*-class subscript for notational ease) can be writen as:

$$
\begin{aligned}
\ln C_{it} \quad = \Big[& \boldsymbol{b} + \sum_{k=1}^{4} \boldsymbol{b}_{yk} \ln y_{kit} + \frac{1}{2}\sum_{k=1}^{4}\sum_{h=1}^{4} \boldsymbol{b}_{ykyh} \ln y_{kit} \ln y_{hit} + \sum_{k=1}^{2} \boldsymbol{b}_{wk} \ln w_{kit} \\
& + \frac{1}{2}\sum_{k=1}^{2}\sum_{h=1}^{2} \boldsymbol{b}_{wkwh} \ln w_{kit} \ln w_{hit} + \sum_{k=1}^{4}\sum_{h=1}^{2} \boldsymbol{b}_{ykwh} \ln y_{kit} \ln w_{hit} \\
& + \boldsymbol{b}_{t} t + \frac{1}{2} \boldsymbol{b}_{tt} t^{2} + \sum_{k=1}^{4} \boldsymbol{b}_{ytk} \ln y_{kit} t + \sum_{k=1}^{2} \boldsymbol{b}_{wk} \ln w_{kit} t \Big] + u_{it} + v_{it}
\end{aligned} \tag{14}
$$

where

$$
u_{it} = \exp\left[\boldsymbol{h}_{1}(t-1) + \cdot \boldsymbol{h}_{2} D_{A} + \boldsymbol{h}_{3} D_{S}\right] \cdot u_{i} \tag{15}
$$

and the latent class probabilities as:

$$
P_{ij}(\boldsymbol{d}_{j}) = \frac{\exp\left[\boldsymbol{d}_{0j} + \boldsymbol{d}_{1j} L_{NBi} + \boldsymbol{d}_{2j} L_{Bi} + \boldsymbol{d}_{3j} D_{NBi} + \boldsymbol{d}_{4j} D_{Bi} + \boldsymbol{d}_{5j} LBR_{i}\right]}{\sum_{j=1}^{J} \exp\left[\boldsymbol{d}_{0j} + \boldsymbol{d}_{1j} L_{NBi} + \boldsymbol{d}_{2j} L_{Bi} + \boldsymbol{d}_{3j} D_{NBi} + \boldsymbol{d}_{4j} D_{Bi} + \boldsymbol{d}_{5j} LBR_{i}\right]} \tag{16}
$$

4. Empirical Results

The SFLCM is estimated with three classes. Likelihood ratio tests have been carried out in order to test whether data supports the model with one class or two classes. Estimating a model with (*J*-1) classes is equivalent to estimating a model with *J* classes but restricting the parameters of the cost frontier equation (*q*) and the latent class probability parameters (*d*) for any two classes to be the same. The likelihood ratio test rejects the one-class model against the model with two classes. Like-wise the two-class model is rejected in favor of the three-classes model.[13] Overall, these tests indicate that no two classes share the same cost frontier and/or the samer error structures. This result suggests that efficiency estimations might be biased if these differences are not controlled. Since our data supports the three-class model, we confine our discussion to the three-class model only.

[13] These Likelihood ratio tests follow a c^2 distribution function with 33 degrees of freedom and the c^2 values are 1889.4 and 254.0, respectively.

To show the advantages of the proposed methodology, we also carried out a traditional cluster analysis to split the sample.[14] The variables considered in the cluster analysis are the same five variables included in the latent class probabilities. In addition, the cluster analysis was carried out using the so-called *k-means* method. This method allow us to work with the same number of classes than in the SFLCM.

Table 2 sumarizes the classifications resulting from both methodologies. Even though they use the same five separating-variables to identify groups of firms, the classifications are differents. The differences are due to the fact that a latent class structure uses not only sample-information (such as a cluster analysis) but also the goodness of fit for each class models.

The main features of banks in each class are summarized in Table 3. The biggest group that is identified using both methodologies (second class) includes most of the banks and it is mainly formed by *commercial banks*. This group includes almost all the savings banks in the sample and a set of regional banks which employ a high proportion of deposits to fund loans to individuals and industrial or commercial firms. A small set of multiple-line or universal banks also belongs to this group. Due to the fact that the largest financial banks in Spain belong to this group, the average size of these banks is much larger than the banks in other classes.

The other two groups are mainly formed by *non-commercial banks* that specialize in activities related to interbanking market. From the results of the cluster analysis, we can identify two different types of non-commercial banks. The first class includes a set of *personal banks* which capture a high proportion of deposits to fund loans to other banks. The third group is formed by *business banks* that are specialized in loans to non-banks supported by deposits from other banks. A joint feature of the non-commercial banks in these two groups is that they usually operate in large population cities with large branches (about 24 workers per branch).

As indicated in Table 2, the non-commercial bank groups obtained from using the SFLCM are quite different from those found using the cluster analysis. This means that significant information *not* contained in sample-separating variables has been used by the SFLCM to split the sample. For instance, the SFLCM classification might reflect

---

[14] See, for instance, Pérez et al. (1999).

unobservable differences in loans (mortgage vs. bussines, short-term vs. long-term) or in borrowed money (demand deposits vs. time or saving deposits) that require different levels of monitoring or different cost structures. If these differences are not controlled for in estimating the technology, the misspecification can be mistakenly identified as inefficiency.

Table 4 reports average cost efficiency estimated using the highest probability cost frontier as a reference technology.[15] This table shows that the average cost efficiency of the Spanish bank sector is 82.1 percent using the SFLCM and 78.0 percent using the traditional cluster analysis. As expected, this result reveals that traditional cluster analyses (that do not take into account unobservable differences in cost structures) tend to *underestimate* the industry average efficiency level. Another interesting feature is the time path of efficiency. The results in Table 4 show a decreasing trend in cost efficiency over the period 1992-2000. The rate of reduction in the cluster analysis is found to be higher compared to the SFLCM. Thus, the cluster analysis, in this application, *overestimate* the rate of change in cost efficiency.

The parameter estimates using the SFLCM are presented in Table 5. The parameter estimates in the cluster analysis are reported in Table 6. In the cluster analysis the cost frontier and efficiency function parameters are estimated separately for each subsample. This is equivalent to estimating a SFLCM but assigning *a priori* a latent class probability equal to one for one class and a probability equal to zero for other classes. This two-stage procedure does not exploit all the information contained in the data because information contained in one class is not used to estimate the parameters in other classes. The results in Tables 5 and 6 reveal that the overall likelihood function value from the SFLCM is clearly higher than that value obtained from the cluster analysis. Therefore, the best model that the data supports is the SFLCM.

We now return to estimation and results other than efficiency. To estimate the cost frontier we normalize all the variables by their respective geometric mean. In this way, the translog cost frontier represents a second-order Taylor approximation, around the

[15] The efficiency levels using (8) are quite similar to those reported in Table 4. This is because estimated posterior probabilities for the highest probability classes are, on average, very high (92.3, 97.8 and 91.1, respectively).

geometric mean, to any generic cost frontier. Since the cost function is homogene of degree one in input prices we need to impose parametric restrictions to ensure that the estimated cost function satisfies linear homogeneity property. In practice, linear homogeneity restrictions are automatically satisfied if the cost and input prices are expressed as a ratio of one input price. Here we use wages (price of labor) as a numeraire. The estimated cost frontier elasticities are found to be positive at the point of approximation, except for output y_4 in the first class in the cluster analysis. The lack of (positive) monotonicity in this output might be due to model misspecifications we mentioned earlier.

The time trend measures variations in cost not explained by other explanatory variables and are usually attributed to exogenous technical change measured by $-\partial \ln C / \partial t$. Thus, a positive sign on it means technical progress (cost diminution over time, ceteris paribus). The results in Table 5 show, in general, a positive technical change, with the exception of the SFLCM technology in the third class. In addition to technical change, the estimated cost frontiers provides a measure of scale economies. Returns to scale can be estimated as one minus the sum of output cost elasticity (RTS = $1 - \sum_k \partial \ln C / \partial \ln y_k$). At the sample mean, this measure is only a function of the first-order output coefficients. The sum of these coefficients is less than unity for all groups of banks indicating the presence of increasing returns to scale. Many of the past banking studies found similar results.

We now examine the behavior of cost efficiency among banks and over time. Except for the third class, we reject the null hypothesis of time-invariant efficiency (i.e. H_0: η_1= 0). That is, the estimated variations in cost efficiency over the period 1992-2000 are significant for most of the Spanish banks. This result suggests that efficiency change should be included in an examination of bank productivity growth. As expected, the sign on the coefficient of D_A is positive and statistically different fron zero in the last two classes. This means that inefficiency increases when acquisition takes place resulting closure of branches, staff relocation, etc. The estimated coefficients on D_S are positive but not statistically different from zero, indicating that savings banks are as efficient as private banks. It is important to note that none of the banks belonging to the first class is involved in acquisition, and all savings banks belong to either class one or class two. In a latent class framework, this does not preclude estimating a coefficient for D_A (D_S) in the first (third) class, as it happens in a cluster analysis (see Table 6). In these cases, the coefficient is estimated by exploiting information of banks belonging to other classes.

Finally, we examine the coefficients of the latent class probabilities. These coefficients are statistically significant thereby indicating that the variables included in the class probabilities do provide useful information in classifying the sample. The sign of these variables suggests that the higher (smaller) the deposits ratio (loans to banks ratio), the smaller (higher) the probability of belonging to the first two classes. The value and the negative sign on the labor to branch ratio in the second class suggest that probability of membership in the second-class decreases significatively when the branch size increases.

5. Conclusions

Estimates of cost efficiency can be biased if firms in an industry use different technologies. In order to reduce the likelihood of misspecification, researches often often classify the sample into groups using sample separation information and then carry out estimation separately on the sub-samples. This procedure, however, is not efficient because it does not take into account inter-class information that can be quite important in most of the empirical applications. In the present paper, we propose using a Stochastic Frontier Latent Class Model that exploit all the information contained in the data efficiently. This model is developed in a "panel data" framework which allows cost efficiency to vary over time in a parametric form.

We include an application of the methodology using Spanish banking data. The classifications resulting from using the SFLCM and a traditional cluster analysis are different, indicating that unobservable differences (in loans, borrowed money, etc., which require different levels of monitoring or different cost structures) are used by the SFLCM to split the sample. The result reveals that the cluster analysis (that do not take into account unobservable differences in cost structures) tends to *underestimate* average efficiency of the banks and *overestimate* the changes in cost efficiency.

References

Battese, G., and T. Coelli (1992), “Frontier Production Functions, Technical Efficiency and Panel Data: with Application to Paddy Farmers in India.” *Journal of Productivity Analysis* 3, 153-69.

Beard, T., S. Caudill, and D. Grooper (1991), “Finite Mixture Estimation of Multiproduct Cost Functions”, *Review of Economics and Statistics* 73, 654-664.

Beard, T., S. Caudill, and D. Grooper (1997), “The Diffusion of Production Processes in the U.S. Banking Industry: A Finite Mixture Approach”, *Journal of Banking and Finance* 21, 721-740.

Caudill, S. (2002), “Estimating a Mixture of Stochastic Frontier Regression Models via the EM Algorithm: A Multiproduct Cost Function Application”, *Empirical Economics,* forthcoming.

Greene, W. (1993), “The Econometric Approach to Efficiency Analysis”, en Fried, H. , Lovell, C.A.K. y Schmidt, S. (eds.), *The Measurement of Productive Efficiency: Techniques and Applications*, Oxford University Press, New York, 68-119.

Greene, W. (2001), “New Developments in the Estimation of Stochastic Frontier Models with Panel Data”, Efficiency Series Paper 6/2001, Departmento de Economica, Universidad de Oviedo. http://www.uniovi.es/eficiencia/

Greene, W. (2002), “Fixed and Random Effects in Nonlinear Models”, Working Paper, Department of Economics, Stern School of Business, NYU.

Jondrow, J., K. Lovell, I. Materov, and P. Schmidt (1982), “On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model”, *Journal of Econometrics* 19, 233-238.

Kolari, J. and A. Zardkoohi (1995), “Economies of Scale and Scope in Commercial Banks with different output Mixes”, Texas A&M Working Paper.

Kumbhakar, S. C. (1990), “Production Frontiers, Panel Data, and Time-Varing Technical Inefficiency”, *Journal of Econometrics*, 46, 201-211.

Lang, G. and P. Welzel (1996), “Efficiency and Technical Progress in Banking: Empirical Results for a Panel of German Co-operative Banks” *Journal of Banking and Finance* 20, 1003-1023.

Lee, Y., and P. Schmidt (1993), “A Production Frontier Model with Flexible Temporal Variation in Technical Efficiency”, en Fried, H., Lovell, C.A.K. y Schmidt, S. (eds.), *The Measurement of Productive Efficiency: Techniques and Applications*, Oxford University Press, Oxford, 3-67.

McAllister, P. and D. McManus (1993), “Resolving the Scale Efficiency Puzzle in Banking”, *Journal of Banking and Finance* 17, 389-405.

Mester, L. (1993), “Efficiency in the Savings and Loan Industry.” *Journal of Banking and Finance* 17, 267-286.

Mester, L. (1997), “Measuring Efficiency at US banks: Accounting for Heterogeneity is Important.” *European Journal of Operational Research* 98, 230-424.

Pérez, F., J. Maudos, and J.M. Pastor (1999), *Sector Bancario Español (1986-1997): cambio estructural and competencia*, Editorial de la Caja de Ahorros del Mediterráneo.

Polachek, S. and B. Yoon (1987), “A Two-tiered Earnings Frontier Estimation of Employer and Employee Information in the Labor Market” *Review of Economics and Statistics* 69, 296-302.

Resti, A. (1997), "Evaluating the Cost-efficiency of the Italian Banking System: What can be Learned from the Joint Application of Parametric and Non-parametric Techniques" *Journal of Banking and Finance* 21, 221-250.

Rogers, K. (1998), "Nontraditional Activities and the Efficiency of U.S. Commercial Banks." *Journal of Banking and Finance* 22, 467-482.

Schmidt, P. and R. Sickles (1984), "Production Frontiers and Panel Data", *Journal of Business and Economic Statistics* 2(4), 367-374.

Sealey, C. and J. Lindley (1977), "Inputs, Outputs, and a Theory of Production and Cost at Depository Financial Institutions." *Journal of Finance* 32(4), 1251-1266.

Tsionas, E. (2000), "Nonnormality in Stochastic Frontier Model with an Application to U.S. Banking", Manuscript, Council of Economic Advisors, Ministry of National Economy, Athens.

Table 1. Cost Frontier Variables: Descriptive Statistics

	Mean	Max	Min	St.Dev.
y_1	1649945	4952474	451	68334229
y_2	1397305	3886538	0	36232525
y_3	2748801	6561083	0	91895929
y_4	43639	117438	0	1431196
w_1	6.060	5.337	0.031	136.960
w_2	87.513	189.338	15.543	3210.884
Costs	276648	718232	65	7129464

Table 2. LCM and Cluster Analysis: Class comparison

		LCM			
		Class 1	Class 2	Class 3	Total
Cluster Analysis	Class 1	0 (0.0) (0.0)	3 (15.8) (2.9)	16 (84.2) (30.2)	19
	Class 2	8 (7.0) (57.1)	95 (83.3) (93.1)	11 (9.6) (20.8)	114
	Class 3	6 (16.7) (42.9)	4 (11.1) (3.9)	26 (72.2) (49.1)	36
	Total	14	102	53	169

Note: Percentages with respect the row (column) total class in the first (second) parenthesis.

Table 3. LCM and cluster analysis: Class features

	LCM					Cluster Analysis				
					Class 1					
Variable	Obs	Mean	Max	Min	St.d.	Obs	Mean	Max	Min	St.d.
L_B	14	12.07	38.95	0.58	12.18	19	60.95	94.36	39.60	14.08
D_B	14	37.58	94.06	1.82	34.14	19	12.51	39.64	0.00	10.84
L_{NB}	14	69.65	97.02	30.07	19.55	19	16.55	36.81	0.14	11.98
D_{NB}	14	48.55	91.48	0.64	33.47	19	66.65	88.12	34.46	16.39
LBR	14	11.23	29.12	3.93	7.29	19	23.89	129.00	3.22	29.05
					Class 2					
Variable	Obs	Mean	Max	Min	St.d.	Obs	Mean	Max	Min	St.d.
L_B	102	21.40	46.90	2.18	8.94	114	20.25	43.92	0.40	9.21
D_B	102	14.01	58.12	1.01	11.56	114	12.86	40.57	1.01	9.13
L_{NB}	102	53.28	74.77	22.66	10.89	114	54.66	90.46	25.50	11.42
D_{NB}	102	73.47	91.07	33.47	12.11	114	73.20	91.48	34.26	11.74
LBR	102	6.69	14.76	3.35	2.47	114	6.84	19.88	3.35	3.01
					Class 3					
Variable	Obs	Mean	Max	Min	St.d.	Obs	Mean	Max	Min	St.d.
L_B	53	41.55	94.36	0.40	24.17	36	30.23	82.90	0.58	22.26
D_B	53	37.30	87.50	0.00	27.49	36	61.90	94.06	23.29	18.28
L_{NB}	53	37.65	96.35	0.14	25.58	36	51.62	97.02	3.69	26.51
D_{NB}	53	38.57	88.12	0.18	27.39	36	16.87	42.07	0.18	13.53
LBR	53	23.50	129.00	2.29	23.47	36	23.65	81.99	2.29	18.84

Note: Balance sheet ratios in percentage. except for the Labor to Branch ratio (LBR).

Table 4. Overall efficiency indexes.

	LCM					Cluster Analysis				
Year	Obs	Mean	Max	Min	St.d.	Obs	Mean	Max	Min	St.d.
92	145	84.3	99.7	42.3	10.8	145	81.6	99.4	42.1	10.6
93	147	83.8	99.6	41.9	11.1	147	80.9	99.4	36.6	10.9
94	144	83.6	100.0	51.1	11.0	144	80.1	99.4	40.5	11.1
95	144	82.7	100.0	50.6	11.5	144	79.1	99.4	46.0	11.6
96	142	82.3	100.0	48.7	11.8	142	78.2	99.4	41.3	12.4
97	139	81.7	100.0	46.8	11.9	139	77.2	99.4	36.6	13.3
98	134	81.0	100.0	44.8	12.3	134	76.2	99.4	31.9	14.2
99	128	79.9	100.0	42.8	12.7	128	75.3	99.3	27.3	15.3
00	122	78.3	100.0	40.8	13.8	122	72.5	99.3	4.1	18.5
All	1245	82.1	100.0	40.8	12.0	1245	78.0	99.4	4.1	13.4

Table 5. LCM Parameter Estimates

	Class 1		Class 2		Class 3	
Parameters	Estimates	Est./s.e.	Estimates	Est./s.e.	Estimates	Est./s.e.
Cost frontier						
$\ln y_1$	0.2074	11.428	0.2019	47.089	0.1761	9.416
$\ln y_2$	0.1131	10.400	0.1798	43.769	0.2805	18.207
$\ln y_3$	0.4698	17.032	0.4587	59.440	0.4471	24.953
$\ln y_4$	0.0667	2.562	0.0841	12.594	0.0183	1.176
$\ln w_1$	0.5879	29.330	0.6262	85.625	0.8349	30.268
$0.5(\ln y_1)^2$	-0.0314	-2.220	0.1100	13.197	0.1182	5.421
$0.5(\ln y_2)^2$	0.0262	6.108	0.0832	17.629	0.0919	10.097
$0.5(\ln y_3)^2$	0.1249	10.373	0.1264	14.984	0.1482	15.997
$0.5(\ln y_4)^2$	0.0575	6.242	0.0395	2.969	0.0064	1.404
$0.5(\ln w_1)^2$	0.1829	15.343	0.0904	7.697	-0.0595	-2.208
$\ln y_1 \cdot \ln y_2$	0.0117	2.080	-0.0347	-6.832	-0.0453	-3.954
$\ln y_1 \cdot \ln y_3$	-0.0393	-3.247	-0.0431	-4.532	-0.0775	-6.020
$\ln y_1 \cdot \ln y_4$	0.0183	3.027	-0.0294	-4.010	0.0132	1.701
$\ln y_1 \cdot \ln w_1$	0.0080	0.874	0.0338	5.096	0.0060	0.328
$\ln y_2 \cdot \ln y_3$	-0.0174	-2.583	-0.0533	-6.405	-0.0711	-8.043
$\ln y_2 \cdot \ln y_4$	-0.0232	-5.531	0.0135	2.366	0.0009	0.179
$\ln y_2 \cdot \ln w_1$	0.0065	0.854	0.0343	7.161	0.0276	2.100
$\ln y_3 \cdot \ln y_4$	-0.0758	-8.753	-0.0315	-2.929	-0.0249	-4.756
$\ln y_3 \cdot \ln w_1$	0.0357	2.600	-0.0127	-1.276	0.0472	3.194
$\ln y_4 \cdot \ln w_1$	-0.0588	-7.111	-0.0207	-2.939	-0.0053	-0.550
t	-0.0394	-6.289	-0.0229	-15.333	-0.0042	-0.486
Intercept	11.199	165.833	11.397	1260.60	10.989	228.56
Efficiency term						
t	0.0550	3.832	0.0568	9.951	0.0103	0.334
D_A	0.0596	1.047	0.0993	3.358	0.8643	3.957
D_S	0.1006	0.252	0.0606	0.434	0.1941	0.064
σ^2	0.1499	1.843	0.0201	4.634	0.1422	2.969
ψ	0.0083	1.708	0.0454	4.223	0.1482	2.436
Probabilities						
Intercept	-7.4647	-0.849	-10.140	-2.201		
L_B	-0.2114	-2.804	-0.1625	-3.312		
D_B	0.1613	1.680	0.2242	3.595		
L_{NB}	-0.0469	-0.920	-0.0830	-2.049		
D_{NB}	0.1874	1.896	0.2663	4.183		
LBR	-0.0393	-1.162	-0.1388	-1.827		
	Observations =	1245		LF =	1451.4086	

Table 6. Cluster Analysis Parameter Estimates

	Class 1		Class 2		Class 3	
Parameters	Estimates	Est./s.e.	Estimates	Est./s.e.	Estimates	Est./s.e.
Cost frontier						
$\ln y_1$	0.2816	10.307	0.2117	30.298	0.1848	6.447
$\ln y_2$	0.4606	12.376	0.2008	35.567	0.2822	10.829
$\ln y_3$	0.1908	13.348	0.5173	45.223	0.4664	16.651
$\ln y_4$	-0.2140	-4.503	0.0770	7.263	0.0100	0.407
$\ln w_1$	0.6930	19.691	0.6611	55.935	0.7890	17.750
$0.5(\ln y_1)^2$	0.1512	9.127	0.1089	8.561	-0.0881	-3.262
$0.5(\ln y_2)^2$	0.1757	7.989	0.0554	12.143	-0.0010	-0.114
$0.5(\ln y_3)^2$	0.0510	12.030	0.1514	18.524	0.0674	3.379
$0.5(\ln y_4)^2$	0.0453	6.154	0.0520	4.211	-0.0016	-0.188
$0.5(\ln w_1)^2$	0.0703	2.149	0.1483	7.810	0.1196	5.802
$\ln y_1 \cdot \ln y_2$	-0.0987	-6.740	-0.0186	-3.044	0.0759	4.253
$\ln y_1 \cdot \ln y_3$	-0.0349	-4.499	-0.0666	-6.799	0.0805	4.125
$\ln y_1 \cdot \ln y_4$	-0.0347	-4.696	-0.0299	-3.059	-0.0065	-0.522
$\ln y_1 \cdot \ln w_1$	0.0019	0.102	0.0093	0.989	-0.0258	-1.112
$\ln y_2 \cdot \ln y_3$	0.0261	2.265	-0.0666	-8.040	-0.1368	-8.389
$\ln y_2 \cdot \ln y_4$	-0.0256	-2.377	0.0131	2.005	0.0234	2.816
$\ln y_2 \cdot \ln w_1$	0.0974	3.914	0.0397	5.781	-0.0201	-1.222
$\ln y_3 \cdot \ln y_4$	-0.0926	-7.128	-0.0407	-4.373	-0.0218	-2.449
$\ln y_3 \cdot \ln w_1$	-0.0414	-2.879	0.0326	3.649	0.0427	1.697
$\ln y_4 \cdot \ln w_1$	-0.0109	-1.079	-0.0532	-5.889	0.0277	1.727
t	-0.0224	-2.486	-0.0143	-5.328	-0.0586	-4.300
Intercept	10.711	165.42	11.301	617.62	11.138	162.05
Efficiency term						
t	0.1561	7.517	0.0216	3.002	0.1286	4.387
D_A	-	-	0.1847	3.967	-0.1281	-0.563
D_S	-	-	-0.1824	-1.302	-	-
σ^2	0.1182	2.571	0.0782	4.942	0.1102	2.975
ψ	0.0350	2.310	0.0438	4.527	0.4461	1.881
Observations	132		884		229	
LF	114.2238		977.6686		-7.5292	
Overall LF			1084.3632			