EFFICIENCY MEASUREMENT USING A LATENT CLASS STOCHASTIC FRONTIER MODEL

Luis Orea
Department of Economics
University of Oviedo
Avda. Del Cristo, s/n 33071 Oviedo, Spain
lorea@correo.uniovi.es

Subal C. Kumbhakar
Department of Economics
State University of New York
Binghamton, NY 13902-6000, USA
kkar@binghamton.edu

March 2003

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 that might be inappropriately labeled as inefficiency if such variations in technology are not taken into account. We address this issue by estimating a latent class stochastic frontier model in a panel data framework. An application of the model is presented using Spanish banking data. Our results show that bank-heterogeneity can be fully controlled when a model with four classes is estimated.

JEL classification No.: C24, C81, D24.

Keywords: Stochastic cost frontier, latent class model, panel data, banks.

1. **Introduction**

Stochastic production (or cost) frontier functions have been 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 a particular industry may use different technologies. In such a case estimating a commom frontier function encompassing every sample observation may not be 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 biased. Furthermore, if the unobserved technological differences are not taken into account in estimation, the effects of these omitted unobserved technological differences 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 the earning frontier functions of employers and employees. 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.[1]

To account for heterogeneity, we advocate using a *single-stage* approach, i.e., a latent class stochastic frontier model (hereafter LCSFM) that combines the stochastic frontier approach and a latent class structure. Recently, a few studies have combined the stochastic frontier approach with the latent class structure in order to estimate a mixture of frontier

[1] It is worth noting that this procedure does not use information contained in one class to estimate the technology 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. Since this kind of information is not exploited, it is possible to say that two-stage procedures are not *efficient*.

functions. In particular, Caudill (2003) introduces an expectation-maximization (EM) algorithm to estimate a mixture of two stochastic cost frontiers in the presence of no sample separation information.[2] Greene (2002) proposes a maximum likelihood LCSFM using sample separation information and allowing for more than two classes.

The main feature of the models proposed by both Caudill and Greene is that they assume independence of the efficiency term over time.[3] This assumption doesn't allow one to test whether the efficiency is time-invariant or not, which is not particularly appealing in a productivity growth study.[4] We avoid this problem by developing a panel data LCSFM in both efficiency and latent class components.

An application of the proposed model is presented using data on the Spanish banking system in which different types of banks coexist. For example, one can distinguish between savings and private banks, which have been regulated differently and have been traditionally specialized in different services. In addition, there are different types of banks within the private banking sector (large multiple-line banks, small regional banks, non-commercial banks that specialize in interbanking activities, etc.). Since our data are not detailed enough to split the sample completely into various types of banks, the latent class model is especially suitable for analyzing the Spanish banking industry. [5]

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

---

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

[3] That is, they model the efficiency terms in a "cross-sectional" framework where a firm observed in two periods is treated as two separate firms. It should be noted, however, that in Greene (2002) the panel structure of the data is not ignored when the latent class part is developed since the class probabilities of observations belonging to the same firm are fixed through time.

[4] These models do not estimate the inefficiency term consistently since its variance does not vanish as the sample size increases. See Schmidt and Sickles (1984) and Greene (1993) for a detailed discussion on this issue.

[5] Sáez, Sánchez and Sastre (1994) show, using confidential data, that savings banks concentrate on retail banking, providing checking, savings and loans service to individuals (especially mortgage loans), whereas regional banks are more involved in commercial and industrial loans. Another difference between these two groups is the fact that savings banks are more specialized than other banks in long-term loans, which do not require continuous monitoring.

2. Panel Data Specification of a Latent Class Stochastic Frontier Model

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

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

where subscripts i = 1,…,N; t =1,…,T_i; and j =1,…,J stand for firm (bank), 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 component $u_{it}|_j$.

Additional structures must 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, σ^2_{vj}. The inefficiency term $u_{it}|_j$ is modeled as the product of a time-invariant firm effect, $u_i|_j$, and a parametric function of time (among other variables), λ_{it}. The term $u_i|_j$ is assumed to come from a non-negative truncated normal distribution with zero mean and variance σ^2_{uj}.

Several forms for the function λ_{it} have been proposed in the literature. A common feature of them is that they are *exclusively* functions of time (i.e. $\lambda_{it}=\lambda_t$). We adopt an exponential form for λ_{it}, but allow other variables that might explain differences over time and/or among firms (e.g., public, private, etc.) to be included in λ_{it}. We specify the inefficiency $u_{it}|_j$ component, in general terms, viz.,

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

where $\eta_j = (\eta_{1j},...,\eta_{Hj})'$ is a $H \times 1$ vector of parameters and $z_{it} = (z_{1it},...,\eta_{Hit})'$ is a $H \times 1$ vector of variables that, in addition to time, might affect inefficiency.[6] This specification

[6] It should be noted that the time variation in this model is deterministic and evolutionary, which might or might not be restrictive.

nests several other parametric functions proposed in the literature as special cases. We get the specification proposed by Battese and Coelli (1992) when $z_{it} = (T-t)$. By specifying $z_{it} = (t, t^2)'$ one gets the specification proposed by Kumbhakar (1990). Finally, the Lee and Schmidt (1993) specification is obtained if z_{it} is a set of T time-dummy variables.

With these distributional assumptions, the log density for firm i, if it belongs to class j, can be written as (see Battese and Coelli, 1992)

$$\begin{aligned}\ln LF_{ij}(\theta_j) = \; & \ln\left[1-\Phi(-z_i^*)\right] + (z_i^*)^2 - \frac{1}{2}\left[\ln 2\pi + \ln\sigma_j^2\right]\cdot T_i - \frac{1}{2}\ln(1-\gamma_j)\cdot(T_i-1) \\ & -\frac{1}{2}\cdot\ln\left[1+\gamma_j\cdot(\textstyle\sum_{i=1}^{T_i}\lambda_{it}(\eta_j)^2-1)\right] - \frac{1}{2}\cdot\sum_{t=1}^{T_i}\left[\varepsilon_{it}(\beta_j)^2/(1-\gamma_j)\sigma_j^2\right]\end{aligned} \tag{3}$$

where

$$z_i^* = \frac{\gamma_j\cdot\sum_{t=1}^{T_i}\lambda_{it}(\eta_j)\cdot\varepsilon_{it}(\beta_j)}{\left\{\gamma_j\cdot(1-\gamma_j)\cdot\sigma_j^2\cdot\left[1+\gamma_j\cdot\left(\sum_{t=1}^{T_i}\lambda_{it}(\eta_j)^2-1\right)\right]\right\}^{1/2}},$$

$\varepsilon_{it} = \varepsilon_{it}(\beta_j) = \ln C_{it} - \ln C\,(y_{it}, w_{it}, t, \beta_j)$; $\sigma_j^2 = \sigma_{vj}^2 + \sigma_{uj}^2$; $\gamma_j = \sigma_{uj}^2/\sigma_j^2$; and $\theta_j = (\beta_j, \sigma_j^2, \gamma_j, \eta_j)$ are the parameters associated with the technology of class j, and $\Phi(\cdot)$ is the standard normal distribution function.[7]

Note that the conditional (on class j) log density in (3) is defined for all the time periods over which firm i is observed, while in Greene (2002) it is defined for firm i at each time t. Thus, the overall contribution of firm i to the conditional likelihood in Greene is obtained as $LF_{ij}(\theta_j) = \Pi_{t=1}^{T_i} LF_{it}(\theta_j)$, where $LF_{it}(\theta_j)$ is the conditional likelihood function for firm i at time t. This, however, cannot be done in our model because firm observations are not independent over time. This is the only difference between our model and the one proposed by Greene (2002). That is, while in Greene the inefficiency term varies freely over time (i.e. $u_{it}|_j$ is *i.i.d.*), the inefficiency function in our specification varies systematically over time in a deterministic fashion (i.e.. $u_{it}|_j = \lambda_{it}(\cdot)\cdot u_i|_j$). The rest of our model (the part associated with the latent class structure) is identical to that of Greene.

[7] For estimation purposes the model above is, however, re-parameterized in terms of $\Psi_j = \sigma_{vj}^2/\sigma_{uj}^2$ which is a useful indicator of the relative importance of noise to inefficiency.

In a latent class model, the unconditional likelihood for firm i is obtained as the weighted sum of their j-class likelihood functions, where the weights are the probabilities of class membership. In this formulation, the probabilities reflect the uncertainty that the researchers might have about the true partitioning in the sample. That is,

$$LF_i(\theta,\delta) = \sum_{j=1}^{J} LF_{ij}(\theta_j) \cdot P_{ij}(\delta_j) \quad , \quad 0 \le P_{ij} \le 1 \quad , \quad \Sigma_j P_{ij} = 1 \tag{4}$$

where $\theta = (\theta_1,...,\theta_J)$, $\delta = (\delta_1,...,\delta_J)$ and the class probabilities are parameterized as a multinomial logit model,

$$P_{ij}(\delta_j) = \frac{\exp(\delta_j' q_i)}{\sum_{j=1}^{J} \exp(\delta_j' q_i)} \quad , \quad j = 1,...,J \quad , \quad \delta_J = 0 \tag{5}$$

where q_i is a vector of firm-specific, but time-invariant, 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 as:

$$\ln LF(\theta,\delta) = \sum_{i=1}^{N} \ln LF_i(\theta,\delta) = \sum_{i=1}^{N} \ln \left\{ \sum_{j=1}^{J} LF_{ij}(\theta_j) \cdot P_{ij}(\delta_j) \right\} \tag{6}$$

Under the maintained assumptions, maximum likelihood techniques will give asymptotically efficient estimates of all the parameters.[8] A *necessary* condition for identifing the parameters of the latent class probabilities is that the sample must be generated from either different technologies or different noise/inefficiency terms. That is, J, the number of classes in equation (6), is taken as given. If J is larger than the "true" number of classes (i.e., if we try to fit a model with "too many" classes) the model will be overspecified and the parameters cannot be estimated.

[8] Note that here both the technology and the probability of particular group membership are estimated simultaneously. Since these class probabilities might be *a priori* nonzero, all the observations in the sample should be used to estimate the underlying technology for each class. In contrast, the standard two-stage procedures implicitly restrict the class probabilities to be equal one for a particular class and zero for the others. This precludes using observations that were allocated to one particular group to estimate other class frontiers.

The estimated parameters can be used to compute the conditional posterior class probabilites. Following the steps outlined in Greene (2002) the posterior class probabilities can be obtained from[9]

$$P(j \mid i) = \frac{LF_{ij}(\theta_j) \cdot P_{ij}(\delta_j)}{\sum_{j=1}^{J} LF_{ij}(\theta_j) \cdot P_{ij}(\delta_j)} \tag{7}$$

This expression shows that the posterior class probabilities depend not only on the estimated δ parameters, but also on the vector θ, i.e., the parameters from the cost frontier. This means that a latent class model classifies the sample into several groups even when sample-separating information is not available. In this case, the latent class structure uses the goodness of fit of each estimated frontier as additional information to identify groups of firms.

In the standard stochastic frontier apporach where the frontier function is the same for every firm, we estimate inefficiency relative to the frontier for all observations, viz, inefficiency from $E(u_{it} \mid \varepsilon_i)$ and efficiency from $E[exp(-u_{it})|\varepsilon_i]$.[10] In the present case, we estimate as many frontiers as the number of classes. What remains an issue here is how to measure the efficiency level of an individual firm when there is no unique technology against which inefficiency is to be computed. There are two ways to solve this problem.

First, we can 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 assigned for that class as its reference technology. Note that this method ignores all other class probabilites although the

[9] It is to be noted that although Greene (2002) 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 pertaining to 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 seem to be important because the expression used here and the one proposed by Greene for estimating $P(j|i)$ are equivalent, except for $LF_{ij}(\theta_j)$ which in Greene 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.

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

(posterior) class probabilities are not zero. This scheme of arbitrary weighting and somewhat *ad hoc* selection of the so-called *reference* technology can be avoided by using the second method, viz.,

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

where $P(j|i)$ is the posterior probability to be in the j^{th} class for a 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 (2002) to get firm-specific estimates of the parameters of the stochastic frontier model. The efficiency results obtained by using (8) would be different from those based on the most likely frontier and using it as the reference technology. The magnitude 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

The LCSFM discussed in the previous section is applied to a panel of Spanish banks observed for the period 1992 to 2000. The number of banks in Spain decreased steadily over the last ten years because of a large number of mergers and acquisitions.[11] Due to a change in the structure of the public balance sheets in 1992 that reduced the amount of information reported by banks, and the fact that the majority of mergers took place in the early 1990s, we use an unbalanced panel of 169 banks for the period 1992-2000.

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

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

proposed by Sealey and Lindley (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 to Spanish banks.[12]

Total cost includes both interest and operating expenses. The interest expenses explain about 71% of total cost and they come 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 rent and occupancy cost, communication expenses, or travel and reallocation 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 are reported in Table 1. All monetary variables were deflated using the GDP deflator index, and are expressed in thousands of Euros (using 2000 as the base year).

Regarding the parametric part of the inefficiency component, we consider three z_{it} variables. The first variable is the time trend (t). Using time only, the specification of $\lambda_{it}(.)$ corresponds to the Battese and Coelli (1992) form. Since $\lambda(.)$ is a function of time with only

[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 fees or non-interest income as an output (for example, Lang and Welzel (1996), Resti (1997) and Rogers (1998)).

one parameter, efficiency either increases, decreases or remains constant. The second variable, D_A, is constructed as follows: It takes a value of zero if the bank doesn't acquire any financial institution, and its value is increased by one every time the bank acquires another bank. Since an acquisition involves structural changes (closure of branches, staff relocation, etc.), we expect an increase (from one period to the next) in inefficiency when an acquisition takes place. The third variable, D_S, is a dummy variables that takes a value of one if the financial institution is a savings banks, and zero otherwise. The coefficient of this variables allow us to test whether savings banks are as efficient as the 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, viz., 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 highly populated cities with large branches.

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

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

where

$$u_{it} = \exp\left[\eta_1(t-1) + \cdot\eta_2 D_A + \eta_3 D_S\right]\cdot u_i \tag{15}$$

Finally, the latent class probabilities are specified as

$$P_{ij}(\delta_j) = \frac{\exp[\delta_{0j} + \delta_{1j}L_{NBi} + \delta_{2j}L_{Bi} + \delta_{3j}D_{NBi} + \delta_{4j}D_{Bi} + \delta_{5j}LBR_i]}{\sum_{j=1}^{J}\exp[\delta_{0j} + \delta_{1j}L_{NBi} + \delta_{2j}L_{Bi} + \delta_{3j}D_{NBi} + \delta_{4j}D_{Bi} + \delta_{5j}LBR_i]} \tag{16}$$

4. Empirical Results

In estimating a latent class model one has to address the problem of determining the number of classes. The AIC and BIC (Schwartz's criterion) are the most widely used in standard latent class models to determine the appropriate number of classes. [13] We have computed AIC and BIC (Schwartz's criterion) statistics in order to select the class size (see, for example, Fraley and Raftery, 1998).[14] Both statistics favor the model's goodness of fit but put a penalty on the number of parameters in the model. Hence, they can be used to compare models with different number of classes. The best model is the one with the lowest AIC or the highest BIC. Table 2 reports the AIC and BIC values. The AIC (BIC) values decrease (increase) as the number of classes increase from one to four, indicating that the preferred model is that with four classes.[15] This result is also in line with the testing 'down' strategy suggested by Greene (2002). The likehood ratio test suggested by Greene rejects models with 3, 2, and 1 classes (testing down from 4 to 3, 3 to 2, and 2 to 1 classes).

We also examine the class selection issue from the efficiency point of view. Since estimated efficiency is likely to be biased if differences in the cost frontier and/or the same error structures are not controlled, one would expect the efficiency levels to increases as the number of classes increases. This is clearly confirmed by the average efficiency scores, shown in Table 3, which are obtained by estimating models with one, two, three and four classes. As expected, the efficiency levels rise as the number of classes increases, indicating that unless bank-heterogeneity is properly taken into account estimated inefficiency is likely to be biased (due to model misspecification) upward. This misspecification is especially serious when a simple model with only one group is estimated. The problem, however, seems

---

[13] These statistics in our latent class framework can be written as:

$$\ln AIC(J) = \ln\left(\sum_{i=1}^{N}\sum_{t=1}^{T_i}\left(\sum_{j=1}^{J} P(j\,|\,i)\cdot \varepsilon_{it}^{2}(j)\right)\right) - \ln\left(\sum_{i=1}^{N} T_i\right) + \frac{2K(J)}{\sum_{i=1}^{N} T_i}$$

$$BIC(J) = 2\cdot \ln LF(J) - K(J)\cdot \ln\left(\sum_{i=1}^{N} T_i\right)$$

[14] The book Applied Latent Class Analysis edited by Hagenaars and McCutcheon (2002) has several applications where the AIC and BIC are used. Many applications in marketing, psychology, social sciences, and other disciplines are cited in there.

[15] We tried to estimate a model with five classes, but failed to achieve convergence. We take this as evidence that a model with five classes is overspecified.

to vanish when a model with four classes is used. Thus, we limit our discussion to the four-class model.

The estimated class probabilities for the highest probability classes and the main features of banks in each class are summarized in Table 4. The posterior class probabilities are, on average, very high (90 percent or more). It is worth noting that the highest values are obtained for the third and fourth classes, where the prior class probabilities are also high. The classification resulting from these probabilities shows that the largest group (third class) is mainly formed by *commercial banks*, which concentrate on retail banking, providing savings and loans services to individuals and loan services to industrial or commercial firms. In particular, this group includes almost all the savings banks in the sample and a set of private banks, formed by regional banks which employ a high proportion of deposits to fund loans and multiple-line or universal banks. The average size of these banks is much larger than the banks in other classes because that the largest financial banks in Spain belong to this group. The other three groups are mainly formed by *non-commercial banks* that specialize in activities related to the interbanking market. A detailed examination of this group would allow us to identify two different types of non-commercial banks. The first type includes a set of *personal banks* which capture a high proportion of deposits to fund loans to other banks. The second type is formed by *business banks* that are specialized in loans to non-banks supported by deposits from other banks. A common feature of all the non-commercial banks is that they usually operate in high population cities with large branches.

Table 5 reports average cost efficiency estimated using the highest probability cost frontier as a reference technology. Since the estimated posterior probabilities for the highest probability classes are very high, the efficiency levels reported in this table are quite similar to those (not shown in Table 5) computed using (8). This table shows that the average cost efficiency of the Spanish banking sector as a whole is 82.8 percent. There are, however, substantial differences in efficiency levels among classes. While the average efficiency in the first class is 88.3%, it decreases to 65.2% in the second class. On the other hand, the efficiency in the largest (third) class is, on average, 86.3%. [16]

[16] In a previous version of this paper, we also carried out a cluster analysis using the variables included in the latent class probabilities as sample-separating information. The classification obtained using this set of variables was quite different from those obtained using the LCSFM. This suggests that the LC methodology uses the

The parameter estimates are presented in Table 6.[17] To estimate the cost frontiers we normalize all the variables by their respective geometric mean. In this way, the translog form represents a second-order Taylor approximation, around the geometric mean, to any generic cost frontier. Since the cost function is homogeneous of degree one in input prices, we need to impose parametric restrictions to ensure that the estimated cost function satisfies this 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. Since all elasticities are positive at the geometric mean, the estimated cost frontiers are increasing in outputs and input prices. The cost frontier should also be concave in input prices. The second derivative of cost frontier with respect to loanable funds price, evaluated at the sample geometric mean, is negative in all cases. Since our cost frontiers include only two inputs and we imposed homogeneity of degree one in input prices, this implies that the Hessian matrix is a negative semidefinite matrix. Therefore, these results confirm (positive) monotonicity of all cost frontiers and indicate that the estimated cost frontiers are concave at the geometric mean.

Variations in cost (over time) that are not explained by other explanatory variables, 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 6 show a technical progress for banks included in classes two and three. No significant shifts in the cost frontier were found in other classes. In addition to technical change, the estimated cost frontiers provide a measure of scale economies. Returns to scale can be

goodness of fit of each estimated frontier to take into account differences (for instance, in mortgage vs. bussines loans, short-term vs. long-term loans, or in demand deposits vs. time or saving deposits) that cannot be controlled by using only sample-separating information. If these differences require different levels of monitoring or different cost structures, and they are not controlled for, the misspecification can be mistakenly identified as inefficiency.

[17] The estimation routines were programmed in Gauss. It is noteworthy that these parameters were obtained by maximizing (6) directly using the BFGS method. The gradient methods might not work in other applications since, in general, estimation of a latent class model is a difficult task. In these situations, the EM algorithm can be used. Further details on this algorithm can be found in Greene (2001).

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. While we cannot reject the null hypothesis of time-invariant efficiency (i.e., H_0: $\eta_1 = 0$) in the first and fourth classes, it can be rejected in the second and third classes. The positive sign on the estimated value of η_1 in these classes indicates a decline in efficiency level of the banks in these classes. Since 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 bank productivity growth studies. As expected, the sign on the coefficient of D_A is positive and statistically different from zero (at the 10% level) in all classes. This means that inefficiency increases when acquisitions take place resulting in closure of branches, staff relocation, etc. The value of this coefficient is larger for banks in the fourth class, which indicates that costs of mergers for banks in this class are higher compared to those in other classes. Finally, the estimated coefficients on D_S are positive but not statistically different from zero, indicating that savings banks are as efficient as private banks.

Finally, we examine the coefficients of the latent class probability functions.[18] In general, these coefficients are statistically significant thereby indicating that the variables included in the class probabilities do provide useful information in classifying the sample. For example, the positive sign on the coefficient of the deposit ratio in the second and third classes suggests that the higher the deposit ratio, the higher is the probability of a bank to belong to these classes. Similarly, the significantly negative value on the labor to branch ratio coefficient in the third (first) class indicates that the probability of membership in these classes decreasess when the branch size increases. Perhaps this explains why banks belonging to the third and first classes operate with small branches (as shown in Table 4). It is also to be

[18] To be precise, if we want to analyze the effects of different variables on the probability of class membership we should focus on the marginal effects instead of the coefficients. Since the sign of the marginal effects depends on the sign of the estimated coefficients and we are not interested in the size of these effects, we have focused only on the estimated coefficients.

noted that none of the coefficients in the first class are significant at the 5% level (although some of them are statistically different from zero at the 10% level).

5. Conclusions

Estimates of cost efficiency based on a single class model are likely to be biased if firms in an industry use different technologies. In order to reduce the likelihood of model misspecification, researchers often classify the sample into groups using sample separation information and then carry out separate estimation on the sub-samples. In the present paper, we propose using a *single-stage* methodology that involves estimating a latent class stochastic frontier model that allows cost efficiency to vary over time in a parametric form. Both efficiency and latent class structures are developed in a panel data framework .

We include an application of the methodology using the Spanish banking data. Bank heterogeneity seems to be controlled when a model with four classes is estimated. This decision is also confirmed by the AIC and BIC. Although there are substantial differences in efficiency levels among classes, the average cost efficiency of the Spanish banking industry is found to be around 82.8 percent. Technical progress is found in models with 2 and 3 classes. These models also show a declining trend in technical efficiency. This finding suggests that efficiency change should be an important factor in bank productivity growth studies. Finally, we find that savings banks are as efficient as private banks, and that structural changes (closure of branches, staff relocation, etc.) resulting from acquisitions processes led to a significant reduction 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. (2003) "Estimating a Mixture of Stochastic Frontier Regression Models via the EM Algorithm: A Multiproduct Cost Function Application", *Empirical Economics,* forthcoming.

Fraley, C. and A.E. Raftery (1998) "How Many Cluster? Which Clustering Method? Answers Via Model-Based Cluster Analysis", Technical Report No. 329, Department of Statistics, University of Washington.

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

Greene, W. (2001) "Fixed and Random Effects in Nonlinear Models", Working Paper, Department of Economics, Stern School of Business, NYU.

Greene, W. (2002) "Alternative Panel Data Estimators for Stochastic Frontier Models", Working Paper, Department of Economics, Stern School of Business, NYU.

Grifell, E. and Lovell, C.A.K. (1997) "The Sources of Productivity Change in Spanish Banking", *European Journal of Operational Research* 98, 364-380.

Hagenaars, J.A. and McCutcheon, A.L. (2002), *Applied Latent Class Analysis*, Cambridge University Analysis, New York, NY.

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”, in Fried, H., Lovell, C.A.K. and Schmidt, S. (eds.), *The Measurement of Productive Efficiency: Techniques and Applications*, Oxford University Press, Oxford, 3-67.

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.

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.

Sáez, F.J., J.M. Sánchez, and T. Sastre (1994) “Los mercados de operaciones bancarias en España: especialización productiva and competencia”. Working Paper 9410, Bank of Spain.

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.

Table 1. Cost Frontier Variables: Descriptive Statistics

	Mean	Max	Min	St.Dev.
y_1	1649945	4952474	451	68334229
y_2	1397305	3886538	1 (1)	36232525
y_3	2748801	6561083	1 (1)	91895929
y_4	43639	117438	1 (2)	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

Notes: (1) this variable took the value 0 in one case, that was replaced by the value 1.
(2) this variable took the value 0 in 12 cases, that were replaced by the value 1.

Table 2. Selection statistics.

No of classes	No of parameters	AIC	BIC
One	27	15.72	676
Two	60	1.72	2462
Three	93	1.22	2615
Four	126	0.86	2707

Table 3. Average efficiency indexes with different number of classes.

Year	Four classes	Three classes	Two classes	One class
1992	84.4	84.3	85.2	68.5
1993	84.1	83.8	84.5	66.9
1994	84.0	83.6	84.1	65.1
1995	83.3	82.7	83.0	63.1
1996	83.2	82.3	82.4	61.6
1997	82.8	81.7	81.6	59.4
1998	82.1	81.1	80.8	57.2
1999	81.3	79.9	79.5	54.7
2000	80.1	78.4	77.6	52.0
All	82.8	82.2	82.0	61.2

Table 4. Prior and posterior class probabilities and class characteristics (Averages in percentage).

Class	Firms	Prior	Posterior	L_B (1)	D_B(1)	L_{NB}(1)	D_{NB}(1)	LBR (2)
1	18	21.1	90.7	34.57	14.35	39.30	65.68	9.18
2	18	26.9	89.6	12.57	32.53	67.54	51.80	12.38
3	92	75.6	96.4	21.44	14.19	53.35	73.30	6.60
4	41	69.2	97.5	39.78	41.96	40.3	35.56	26.17

Notes: (1) Balance sheet ratios; (2) Labor to Branch Ratio.

Table 5. Average efficiency indexes.

	Overall sample		Class 1		Class 2		Class 3		Class 4	
Year	Mean	Obs.	Mean	Obs.	Mean	Obs.	Mean	Obs.	Mean	Obs.
1992	84.4	145	86.1	16	69.7	14	88.8	85	77.7	30
1993	84.1	147	86.2	16	70.9	16	88.3	83	78.8	32
1994	84.0	144	88.1	15	69.8	15	87.7	80	79.7	34
1995	83.3	144	88.4	15	67.2	16	86.9	77	80.5	36
1996	83.2	142	88.6	15	65.9	16	86.3	75	82.1	36
1997	82.8	139	88.8	15	64.6	16	85.4	74	83.1	34
1998	82.1	134	89.5	14	63.2	16	84.9	72	82.1	32
1999	81.3	128	89.5	14	59.4	15	84.1	69	81.9	30
2000	80.1	122	90.2	14	56.5	16	83.2	62	81.5	30
All	82.8	1245	88.3	134	65.2	140	86.3	677	80.8	294

Table 6. LCM Parameter Estimates

	Class 1		Class 2		Class 3		Class 4	
Parameters	Estimates	Est./s.e.	Estimates	Est./s.e.	Estimates	Est./s.e.	Estimates	Est./s.e.
Cost frontier								
$\ln y_1$	0.218	15.841	0.206	11.603	0.198	43.531	0.158	8.016
$\ln y_2$	0.218	17.831	0.108	10.213	0.185	47.703	0.292	19.708
$\ln y_3$	0.508	25.799	0.482	22.279	0.452	56.068	0.467	25.692
$\ln y_4$	0.010	0.383	0.103	4.354	0.093	13.961	0.015	0.946
$\ln w_1$	0.663	26.718	0.573	34.819	0.637	89.008	0.843	29.394
$0.5(\ln y_1)^2$	0.112	5.465	-0.024	-1.744	0.105	11.886	0.124	5.668
$0.5(\ln y_2)^2$	0.137	11.822	0.025	6.235	0.086	20.709	0.088	8.734
$0.5(\ln y_3)^2$	0.169	10.889	0.144	11.721	0.122	11.258	0.154	16.198
$0.5(\ln y_4)^2$	0.058	8.703	0.068	7.780	0.043	3.160	0.005	1.100
$0.5(\ln w_1)^2$	0.160	5.168	0.170	15.184	0.078	6.262	-0.071	-2.585
$\ln y_1 \cdot \ln y_2$	-0.051	-3.346	0.013	2.610	-0.034	-7.299	-0.041	-3.499
$\ln y_1 \cdot \ln y_3$	-0.059	-3.320	-0.044	-4.186	-0.035	-3.112	-0.083	-6.519
$\ln y_1 \cdot \ln y_4$	-0.018	-1.886	0.018	3.336	-0.030	-4.107	0.008	0.951
$\ln y_1 \cdot \ln w_1$	-0.009	-0.679	0.005	0.571	0.044	5.861	0.002	0.080
$\ln y_2 \cdot \ln y_3$	-0.064	-5.593	-0.022	-4.417	-0.055	-6.619	-0.073	-8.408
$\ln y_2 \cdot \ln y_4$	-0.067	-9.945	-0.022	-5.890	0.014	2.516	0.005	1.078
$\ln y_2 \cdot \ln w_1$	0.137	14.575	-0.005	-0.658	0.031	6.599	0.031	2.047
$\ln y_3 \cdot \ln y_4$	-0.043	-5.890	-0.084	-9.418	-0.035	-3.073	-0.023	-3.842
$\ln y_3 \cdot \ln w_1$	-0.001	-0.087	0.030	2.684	-0.016	-1.105	0.040	2.624
$\ln y_4 \cdot \ln w_1$	-0.055	-7.605	-0.048	-6.452	-0.028	-3.336	0.006	0.565
t	0.009	1.751	-0.042	-7.226	-0.022	-14.695	0.005	0.502
Intercept	11.304	349.898	11.232	188.175	11.394	1264.375	10.951	210.661
Efficiency term								
t	-0.023	-1.177	0.050	4.641	0.058	9.864	-0.036	-1.069
D_A	0.147	1.788	0.213	3.543	0.137	4.355	1.165	4.623
D_S	0.224	0.513	0.043	0.297	0.074	0.410	0.576	0.086
σ^2	0.028	2.156	0.156	2.160	0.018	4.181	0.156	2.826
ψ	0.044	1.982	0.007	1.975	0.043	3.849	0.134	2.345
Probabilities								
Intercept	5.987	1.342	-2.076	-0.434	-3.631	-0.821	0.000	
L_B	-0.076	-1.826	-0.191	-2.748	-0.129	-2.947	0.000	
D_B	-0.033	-0.807	0.085	1.906	0.140	2.944	0.000	
L_{NB}	-0.075	-1.757	-0.040	-0.862	-0.075	-1.931	0.000	
D_{NB}	0.025	0.860	0.113	2.578	0.176	3.897	0.000	
LBR	-0.114	-1.707	-0.025	-0.850	-0.177	-2.213	0.000	

Number of observations = 1245
Log-likelihood= 1548.7302