A Probabilistic Approach for Assessing the Significance of Contextual Variables in Nonparametric Frontier Models: An Application to Brazilian Banks[*]

Geraldo da Silva e Souza[**]
Roberta Blass Staub[***]
Benjamin Miranda Tabak[****]

Abstract

This article presents an empirical banking application illustrating the use of a nonparametric frontier model relying on a probabilistic definition of the production frontier. It investigates the statistical significance of nonperforming loans in productive efficiency for a sample of Brazilian banks using the concepts of conditional and unconditional efficiency measures in a context where it is not necessary to impose any particular distribution on the production data. The analysis is robust relative to the assumptions of convexity and separability of the underlying technology. It is concluded that, on average, a relative increase of 1% in nonperforming loans implies a statistically significant relative reduction of 5.7% in the conditional measure of efficiency.

Keywords: Probabilistic Frontier, Data Envelopment Analysis, Contextual Variables, FDH, FDH Conditional.

JEL Codes: C14, G21, G33.

1. Introduction

The two-stage data envelopment analysis (DEA) is the typical approach to assess the statistical significance of contextual variables in efficiency analysis. Firstly, one obtains a DEA efficiency measurement and, in a second step, one regresses the estimated efficiency measurement on the covariates of interest, assuming an underlying parametric model. Recent examples of this procedure are Banker and Natarajan (2004, 2008), Souza et al. (2006), Hoff (2006) and Souza and Staub (2007). The validity of the analysis relies on the assumptions of convexity and separability of the underlying technology. Separability means that the production frontier does not change with distinct values of the contextual variables. To relax these assumptions and to provide a more natural data generating process to assess the influence of contextual variables, an alternative method of analysis has been developed by Daraio and Simar (2005, 2007), which is known as the probabilistic approach. The probabilistic approach is intended to measure the distortion caused by the covariate on the frontier. The idea is to define a new concept of production frontier and base the notion of efficiency measurement on probability distributions.

The stochastic characterization of the frontier and the corresponding efficiency measure allow assessing the influence of contextual variables on the production process by the statistical analysis of the ratio conditional to unconditional measures of efficiency. We contribute to the literature on this subject using this new methodology in an empirical application to Brazilian banks.

Our main interest is to assess, in a sample, the significance of nonperforming loans on the probability-based measures of efficiency.

The importance of controlling for nonperforming loans in the assessment of bank efficiency has been extensively discussed in Bernstein (1996) and in Das and Ghosh (2006). These articles show that it is crucial to include nonperforming loans as an environmental variable that explains bank efficiency. Additionally to nonperforming loans, we also study the significance of other covariates such as size, ownership and bank specialization.

The probability-based measures of efficiency rely on a stochastic formulation of the production process. This is described by the joint probability measure of inputs and outputs. The support of this distribution is the attainable production set. The notion of conditional distribution is used to define the frontier. The inclusion of contextual variables is done by conditioning the joint distribution on the vector of contextual variables.

The remainder of our article is structured as follows. Section 2 describes the production banking system we use in our application and the contextual variables of interest. Section 3 presents the probabilistic frontier approach. Section 4 deals with estimation of probability-based measures of efficiency. Section 5 extends the concept of probability-based efficiency measurements to allow contextual variables in the analysis. Section 6 deals with statistical inference and modeling. Finally, in Section 7, we summarize the empirical results and the conclusions of our study.

2.   Production Systems in Banking

The definition of outputs and inputs in banking is controversial. See Colwell and Davis (1992), Berger and Humphrey (2000) and Campos (2002) for an in-depth discussion on the matter. As described in Campos (2002), basically, two approaches are possible – production and intermediation. The production approach considers banks as producers of deposits and loans using capital and labor as inputs. In such a context, output is measured by the number of deposit accounts and the number of transactions performed. Under the intermediation approach, banks function as financial intermediaries converting and transferring financial assets between surplus units and deficit units. Each output is measured in value, not in number of transactions or accounts.

There is not a unique recommendation on what should be considered as the proper set of inputs and outputs, particularly under the intermediation approach. Here we follow the intermediation approach and take the vector defined by the variables securities, loans and demand deposits as output, combined into a single measure (total assets), representing the sum of their values. Combining output is not a strange concept in the intermediation approach. Examples of this practice are given by Hanes and Thompson (1999), Nakane and Weitraub (2005), and Souza et al. (2006). It has the advantage of avoiding spurious efficiency measurements resulting from unique bank specializations.

Although the definition of outputs above is not always followed in the banking literature, it is the most common one, as seen in Campos (2002). Notice, for example, that the usage of demand deposits in the Brazilian banking literature also varies. Nakane (1999) studying cost efficiency considers it as a covariate in the cost function although its specification in the translog cost function is similar to an output. Silva and Jorge-Neto (2002), also working with cost functions, consider demand deposits only as a factor influencing the technical efficiency component in the model.

All production variables are measured as indices relative to a benchmark and are normalized by a measure of size. This approach has the advantage of making the banks more comparable through the reduction of variability and of the influence of size in the efficiency analysis.

The inputs we consider are labor, the stock of physical capital, which includes the book value of premises, equipment, rented premises and equipment and other fixed assets, and loanable funds, which include transaction deposits and purchased funds.

Typically, the output-oriented efficiency analysis is specified using input and output measured in physical quantities. This is not strictly necessary and does not prevent its use in the intermediation approach. One may work with indexes or proxies reflecting the intensity of usage of each variable (input or output) in the production process. This is the case with the present application. Total output, loanable funds and capital are values. Also, we found labor costs to be a more

reliable measure of the intensity of labor usage than the number of employees, which was highly variable within the year. In this context, we defined indexes to reflect the behavior of the production variables.

The database used is COSIF (Financial Institutions Accounting Plan), the plan of accounts comprising balance sheet and income statement items that all Brazilian financial institutions have to report to the Central Bank on a monthly basis. This is the database mostly used in studies dealing with Brazilian banking. See, for example, Nakane (1999) and Campos (2002). The total number of banks used in our analysis (the sample size) is 94.

The statistical analysis carried out in this article is for the year 2001. The year choice allows comparisons with similar studies based on the deterministic DEA and using the same data. See Souza et al. (2006).

The benchmark for each variable, whether an input, an output or a continuous covariate, was chosen to be the median value for the variable. Banks with a value of zero for one of the inputs or the outputs were eliminated from the analysis.

Outputs, inputs, and the continuous covariate were further normalized through the division of their respective indexes by an index of personnel intended to be a size-adjusting factor. The construction of this index follows the same method used for the other variables, that is, the corresponding index is the ratio of the number of employees in December of 2001 by its median value in the same month.

Even after size adjustments some banks still show out-of-range values either for inputs or outputs, since there are outliers in the data. This is a problem in efficiency analysis, which is known to be very sensitive to outliers. To eliminate nonconforming output and input vectors we use a sort of Mahalanobis distance of common use in regression analysis to identify outlying observations. See Souza (1998). This amounts to identifying as outlying observations those for which the ith element of the diagonal of the projection matrix $W(W'W)^{-1}W'$ is at least two times its trace. Here $W = (1, Y)$ or $W = (1, X)$ where 1 is a column of ones and Y and X are the matrices of output products and input usage, respectively.

The covariate of interest for our analysis – a factor likely to affect inefficiency – is nonperforming loans. The importance of nonperforming loans has been subject of an extensive debate in the banking literature. The work of Bernstein (1996) shows that nonperforming loans affect both the level of bank costs and estimates of economies of scale in banking, and that the cost curve estimated for banks with low nonperforming ratios suggests that economies of scale may exist even for the largest banks. Das and Ghosh (2006) study the case of Indian banks and show that, irrespective of the choice of inputs and outputs, high levels of nonperforming loans are associated with low efficiency estimates and vice versa.

Berger and DeYoung (1997) employ a Granger causality test, assuming an intertemporal relation between efficiency (cost and/or profit), capitalization and loan loss provisions of the bank. Within this approach, they test for different hypotheses of management behavior:

1. the bad management hypothesis, in which low cost efficiency Granger causes high loan loss;

2. the skimping hypothesis, in which high cost efficiency Granger causes high loan loss;

3. the bad luck hypothesis, in which exogenous shocks in the level of loan loss provision Granger cause changes in cost efficiency levels; and

4. the moral hazard hypothesis, in which thinly capitalized banks are more prone to risk-taking.

These hypotheses have been tested for a variety of countries and banking systems. Berger and DeYoung (1997) find evidence of bidirectional causality between nonperforming loans and bank cost efficiency for the U.S. Williams (2004) studies European savings banks and finds evidence supporting the bad management hypothesis. On the other hand, Rossi et al. (2005) investigate banks from transition countries and find support only to the bad luck hypothesis.

In the Brazilian context nonperforming loans may be very important, as the development of credit risk models is relatively new in the banking system. Scarce resources in the analysis of credits granted by banking institutions imply inefficient monitoring of the quality of loans.

3. Probability-Based Efficiency Measurements

In this section, we present a probabilistic interpretation of the Farrell-Debreu efficiency scores (Coelli et al., 2005) which provides a new way to describe the nonparametric free disposal hull (FDH) estimator. The formulation was proposed by Daraio and Simar (2005).

Let $\Psi = \left\{ (x, y) \in \mathrm{R}_+^{p+s}, x \text{ can produce } y \right\}$ and let (X, Y) denote a joint random variable with values in Ψ defining the production process, i.e., generating the input-output observations. Notice that input is p dimensional and output is s dimensional.

The probability approach to efficiency is based on the following function defined on Ψ:

$$H_{XY}(x, y) = P(Y \geq y, X \leq x)$$

We notice that

1. $H_{XY}(y, x)$ gives the probability that a unit operating at input, output levels (x, y) is dominated, i.e., that another unit produces at least as much output while using no more of any input than the unit operating at (x, y);

2. $H_{XY}(y, x)$ is monotone nondecreasing in x for each y and monotone nonincreasing in y for each x.

Applying Bayes' rule to the probability function $H_{XY}(x, y)$ we get

$$H_{XY}(x, y) = P(X \le x | Y \ge y)P(Y \ge y) = F_{X/Y}(x|y)S_Y(y)$$

and

$$H_{XY}(x, y) = P(Y \ge y | X \le x)P(X \le x) = S_{Y/X}(y|x)F_X(x)$$

The concepts of probability-based measures of efficiency are defined for the input-oriented and output-oriented cases, assuming $S_Y(y) > 0$ and $F_X(x) > 0$, respectively. For input orientation

$$\theta(x, y) = inf\left\{\theta | F_{X/Y}(\theta x | y) > 0\right\} = inf\left\{\theta | H_{XY}(\theta x, y) > 0\right\}$$

and for output orientation

$$\lambda(x, y) = sup\left\{\lambda | S_{Y/X}(\lambda y | x) > 0\right\} = sup\left\{\lambda | H_{XY}(\lambda y, x) > 0\right\}$$

The two measurements allow the following production interpretations:

1. Input case: $\theta(x, y)$ is the highest proportionate reduction of inputs required for a unit operating at (x, y) to achieve zero probability of being dominated, holding the output level fixed. The frontier production pair in this case is given by $(\theta(x, y)x, y)$;

2. Output case: $\lambda(x, y)$ is the highest proportionate increase in outputs required for a unit operating at (x, y) to achieve zero probability of being dominated, holding the input level fixed. The frontier production in this case is given by $(x, \lambda(x, y)y)$.

In our empirical work we choose output orientation since this orientation provides a comparison with previous studies using DEA, see Souza et al. (2006). In this context, we will focus our discussion on the output case.

4. Estimating the Output-Oriented Probability-Based Measure of Efficiency

Consider the n input-output observations $(x_\nu, y_\nu) \in \Psi$. The output efficiency measure $\lambda(x_\nu, y_\nu)$ is estimated by

$$\hat{\lambda}(x_\nu, y_\nu) = sup\left\{\lambda | \hat{S}_{Y|X}(\lambda y_\nu | x_\nu) > 0\right\}$$

with

$$\hat{S}_{Y|X}(y_\nu|x_\nu) = \frac{\sum_{j=1}^{n} I(x_j \le x_\nu, y_j \ge y_\nu)}{\sum_{j=1}^{n} I(x_j \le x_\nu)}$$

where the inequalities are element-by-element.

It can be shown that

$$\hat{\lambda}(x_\nu, y_\nu) = \max_{j|x_j \le x_\nu} \left\{ \min_{l=1,\ldots,s} (\frac{y_j^l}{y_\nu^l}) \right\}$$

The quantity y_j^l is the lth element of vector y_j.

As noted by Cazals et al. (2002), $\hat{\lambda}(x_\nu, y_\nu)$ coincides with the FDH estimator.

The FDH estimator is a consistent estimator of $\lambda(x, y)$. For a convex technology both FDH and DEA will be consistent with rates of convergence $n^{-\frac{2}{p+s+1}}$ and $n^{-\frac{1}{p+s}}$, respectively. The rate of convergence of the FDH is slower, but this estimator is robust relative to the convexity assumption.

5.   Assessing the Significance of Contextual Variables

We now consider the production process as being generated by the joint distribution of (X, Y, Z). This is a probability measure defined on $\mathrm{R}_+^p \times \mathrm{R}_+^s \times \mathrm{R}^u$. This probability distribution generates the n production observations (x_ν, y_ν, z_ν). The z_ν are realizations of the vector Z of contextual variables.

The attainable production set when $Z = z$ is defined by the support of

$$
\begin{aligned}
H_{XY}(x, y|z) &= P(X \le x, Y \ge y|Z = z) \\
&= P(Y \ge y|X \le x, Z = z)P(X \ge x|z = z) \\
&= S_{Y|X,Z}(y|x, z)F_{X|Z}(x|z)
\end{aligned}
$$

for all x such that $F_{X|Z}(x|z) > 0$.

The output-oriented measure of efficiency conditional on $Z = z$, is defined by

$$\lambda(x, y|z) = sup \left\{ \lambda|S_{Y|X,Z}(\lambda y|x, z) > 0 \right\} \tag{1}$$

Assume that covariate Z is absolutely continuous and let $K(.)$ be a kernel with compact support.[1] We take the quantity

$$\hat{\lambda}(x_\nu, y_\nu)|z_\nu = \frac{\sum_{j=1}^{n} I(x_j \le x_\nu, y_j \ge y_\nu)K(\frac{z_\nu - z_j}{h})}{\sum_{j=1}^{n} I(x_j \le x_\nu)K(\frac{z_\nu - z_j}{h})}$$

---

[1] For kernels with unbounded support the conditional measure of efficiency coincides with the FDH estimator.

as an estimate of $\lambda(x_\nu, y_\nu | z_\nu)$ where h is the bandwidth corresponding to kernel $K(.)$. There are alternative ways to choose h. Here we minimize the approximated mean integrated square error. The technique is available in SAS (2007). See Silverman (1986) for more details.

For multivariate covariates, joint product kernels seem to be easier to handle. In this case h can be defined by a vector formed by independent marginal choices of bandwidths.

If the covariate is categorical or a mixture of continuous and categorical variables, enough replications of the levels of categorical variables will be necessary to estimate $\lambda(x, y | z)$.

One can show that

$$\hat{\lambda}(x_\nu, y_\nu | z_\nu) = \max_{j | x_j \leq x_\nu, |z_j - z_\nu| < h} \left\{ \min_{l=1,\ldots,s} \left(\frac{y_j^l}{y_\nu^l} \right) \right\}$$

The analysis of the effect of Z on the measure of efficiency (FDH) is performed considering the ratios

$$R_\nu = \frac{\hat{\lambda}(x_\nu, y_\nu | z_\nu)}{\hat{\lambda}(x_\nu, y_\nu)}$$

If the ratio increases with the covariate, we see evidence of a favorable Z. In this context, Z can be regarded as an extra freely available input. On the other hand, if the ratio decreases with the covariate, we see evidence of an unfavorable Z. In this case the covariate works as a compulsory or unavoidable output to be produced in face of a negative environmental condition induced by Z.

6. Empirical Results

We now consider the problem of assessing the statistical significance of non-performing loans to some measure of efficiency associated with the production frontier of a sample of 94 Brazilian banks. The analysis is of relevance since a positive association would point to the use of efficiency measurements as an additional indicator of potential bank failure. On the other hand, bank efficiency is also an important aspect in global market competition. Overall, the identification of characteristics influencing bank efficiency allows a better understanding of the banking system, resulting in social and economic benefits.

Souza et al. (2006) investigated Brazilian bank efficiencies using two-stage DEA Tobit models. They report a negative effect of nonperforming loans which, however, is not statistically significant. For the same data we refine their study considering the effect of nonperforming loans on the FDH frontier.[2]

---

[2]The Matlab routines to compute the conditional and unconditional probabilistic measures of efficiency where gently provided by Leopold Simar and Cinzia Daraio.

Let q represent nonperforming loans. We begin our analysis by comparing the estimates $\hat{\lambda}(x_\nu, y_\nu)$ and $\hat{\lambda}(x_\nu, y_\nu | q_\nu)$. Significant differences in the distributions of these numbers indicate the influence of q.

The value of h that minimizes the approximated mean integrated error for the quadratic kernel is 0.5308. The Spearman rank correlation between the values R_ν and q_ν is -0.32 with a p value of 0.002, indicating a significant negative association. The two-sample Kolmogorov-Smirnov test also rejects the hypothesis of equality of the distributions at the 1% level. The sample mean of the R_ν is 0.902 with a standard error of 0.022 and it is significantly lower than one providing further evidence of the negative effect of q.

To investigate the association between the frontier and q in more detail, and motivated by the work of Souza and Staub (2007), we fit distinct gamma distributions for the quantities R_ν. The gamma family we use has a constant shape parameter P with scale parameters $\omega_\nu = \exp(-\mu_\nu)$ where $\mu_\nu = \beta_0 + z_{1\nu}\beta_1 0 + \ldots + z_{u\nu}\beta_u$. The $z_{k\nu}$ are realizations of the covariates $z = (z_1, \ldots, z_u)$ and the β_k are parameters to be estimated. In our application, the vector z has dimension 11 and is composed of the categorical variables bank nature (n), bank type (t), bank size (s), bank control (c) and bank origin (o), in addition to nonperforming loans (q). The variable n assumes one of three values (commercial, multiple, other) and is represented by two dummy variables n_1 and n_2 corresponding to commercial and multiple types. The variable t assumes one of four values (credit, business, bursary, retail) and is represented by three dummy variables t_1, t_2, and t_3. The omitted dummy stands for retail. The variable s assumes one of four values (large, medium, small, micro), and is also represented by three dummy variables s_1, s_2, and s_3. The omitted dummy stands for micro. The variable c assumes one of two values (private, public), and is represented by c_1, the dummy for private. Finally, the variable o assumes one of two values (domestic, foreign) and is represented by the dummy o_1 for domestic.

Descriptive statistics for the entire population for all production related variables are shown in Table 1. Regarding the statistics for categorical variables, Table 2 shows summary measures for the ratio of conditional to unconditional measures of efficiency. In Table 3, we provide the definitions of all continuous variables used in this article as a function of the COSIF accounting system. The accounts are spelled out in Table 4. The number of employees and the categorical variables were provided by the Central Bank of Brazil.

Table 1

Descriptive statistics for output, inputs (cost of labor, stock of physical capital and loanable funds), $\hat{\lambda}_n(x, y|q)$ and $\hat{\lambda}_n(x, y)$

Variable	N	Mean	Std Dev	Median	Minimum	Maximum
Output	94	0.932	1.126	0.459	0.033	4.911
Cost of labor	94	0.817	0.442	0.726	0.148	2.574
Stock of physical capital	94	1.004	0.882	0.791	0.013	5.886
Loanable funds	94	1.041	1.542	0.369	0.016	8.455
Nonperforming loans (q)	94	0.897	1.534	0.394	0.000	12.190
$\hat{\lambda}_n(x, y|q)$	94	1.525	1.210	1.000	1.000	7.899
$\hat{\lambda}_n(x, y)$	94	1.926	1.870	1.125	1.000	13.295

Table 2

Descriptive statistics for categorical variables. Response is $\frac{\hat{\lambda}_n(x, y|q)}{\hat{\lambda}_n(x, y)}$

Variable	Level	N	Mean	Std Dev	Minimum	Maximum
Bank nature	Comercial (n_1)	12	0.951	0.169	0.413	1
	Multiplo (n_2)	81	0.893	0.226	0.201	1
Bank type	Credit (t_1)	33	0.871	0.258	0.225	1
	Business (t_2)	24	0.873	0.212	0.413	1
	Bursary (t_3)	3	0.763	0.411	0.289	1
	Retail (t_4)	34	0.964	0.144	0.201	1
Bank size	Large (s_1)	18	0.968	0.096	0.636	1
	Medium (s_2)	30	0.901	0.203	0.201	1
	Small (s_3)	25	0.822	0.294	0.225	1
	Micro (s_4)	21	0.939	0.192	0.324	1
Bank control	Private (c_1)	79	0.886	0.234	0.201	1
	Public (c_2)	15	0.986	0.056	0.784	1
Bank origin	Foreign (o_1)	28	0.861	0.244	0.289	1
	Domestic (o_2)	66	0.919	0.206	0.201	1

Table 3

COSIF account compositions used to build the variables output, inputs (cost of labor, stock of physical capital and loanable funds), nonperforming loans

Variable	COSIF accounts composition
Output	1.3.1.00.00-7+1.3.2.00.00-0+1.6.0.00.00-1+(-1.6.9.00.00-8) +4.1.1.00.00-0
Cost of labor	8.1.7.27.00-3+8.1.7.30.00-7+8.1.7.33.00-4+8.1.7.36.00-1
Stock of physical capital	2.2.3.00.00-1+2.2.4.00.00-4+2.2.9.00.00-9
Loanable funds	4.1.2.00.00-3+4.1.3.00.00-6+4.1.5.00.00-2+4.1.8.00.00-1 +4.1.4.00.00-9+4.1.6.00.00-5+4.1.7.00.00-8+4.2.0.00.00-6
Nonperforming loans	+4.3.0.00.00-5+4.6.0.00.00-2+1.6.9.00.00-8

Table 4
COSIF accounts

Account number	Description
1.3.1.00.00-7	Free (bonds, stocks and derivatives)
1.3.2.00.00-0	Linked to repo operations
1.6.0.00.00-1	Credit
1.6.9.00.00-8	Provisioning
1.6.9.00.00-8	Nonperforming loans
2.2.3.00.00-1	Owner-occupied land and building
2.2.4.00.00-4	Installations and furniture
2.2.9.00.00-9	Others
4.1.1.00.00-0	Deposits
4.1.2.00.00-3	Savings deposits
4.1.3.00.00-6	Interfinancial deposits
4.1.4.00.00-9	Deposits under warning
4.1.5.00.00-2	Time deposits
4.1.6.00.00-5	Liabilities from special deposits and funds and programs
4.1.7.00.00-8	APE - Special deposits
4.1.8.00.00-1	Foreign exchange deposits
4.2.0.00.00-6	Repo liabilities
4.3.0.00.00-5	Resources from FX instruments, notes, mortgages, bonds and similar instruments
4.6.0.00.00-2	Liabilities from loans
8.1.7.27.00-3	Personnel expenses - benefits
8.1.7.30.00-7	Personnel expenses - social insurance
8.1.7.33.00-4	Personnel expenses - remuneration
8.1.7.36.00-1	Personnel expenses - training

We write,

$$
\begin{aligned}
\mu_\nu &= \beta_0 + \beta_1 n_{1\nu} + \beta_2 n_{2\nu} + \beta_3 t_{1\nu} + \beta_4 t_{2\nu} + \beta_5 t_{3\nu} \\
&+ \beta_6 s_{1\nu} + \beta_7 s_{2\nu} + \beta_8 s_{3\nu} + \beta_9 c_{1\nu} + \beta_{10} o_{1\nu} + \beta_{11} q_\nu
\end{aligned}
$$

Since qualitative factors are present and not enough replications are available we assume separability of these factors.

Table 5 shows the statistical results from maximum likelihood estimation. There is no evidence that the categorical variables are statistically significant. Contrary to the DEA findings reported by Souza et al. (2006), where q has a negative coefficient but is not significant, one can infer that the conditional efficiency ratio is significantly and negatively related to nonperforming loans, which suggests that the latter can be interpreted as an undesired output. Nonperforming loans require the use of more inputs in the production activity, perhaps due to the fact that banks have to monitor these loans and bear the associated costs. Differences relative to the DEA approach also suggest a non-convex or non-separable technology relative to nonperforming loans.

Table 5

Parametric model for the ratio $\frac{\hat{\lambda}_n(x,y|q)}{\hat{\lambda}_n(x,y)}$, assuming a gamma distribution with shape parameter P and scale parameters $\exp(-\mu_\nu)$ where μ_ν is a linear function of $n_1, n_2, t_1, t_2, t_3, s_1, s_2, s_3, c_1, o_1$, and q

Var	Estimate	Standard error	t	p-value	Lower	Upper
Intercept	-2.545	0.365	-6.96	$<.0001$	-3.271	-1.820
n_1	0.070	0.319	0.22	0.827	-0.564	0.703
n_2	0.070	0.309	0.23	0.822	-0.544	0.683
t_1	-0.001	0.122	-0.01	0.992	-0.242	0.240
t_2	-0.050	0.098	-0.51	0.611	-0.245	0.145
t_3	-0.179	0.193	-0.93	0.356	-0.562	0.204
s_1	0.026	0.143	0.18	0.854	-0.257	0.310
s_2	-0.027	0.108	-0.25	0.801	-0.242	0.187
s_3	-0.093	0.097	-0.96	0.340	-0.287	0.100
c_1	-0.074	0.105	-0.71	0.481	-0.283	0.134
o_1	0.056	0.072	0.77	0.442	-0.088	0.199
q	-0.064	0.021	-3.01	0.003	-0.105	-0.022
P	12.098	1.741	6.95	$<.0001$	8.641	15.554

Neglecting the categorical contextual variables and conditional on the sample mean value $\bar{q}$ of nonperforming loans, the elasticity of the conditional measure of efficiency is estimated by $\bar{q}\hat{\beta}_{11} = 0.897(-0.064) = -0.057$ with a standard error of 0.019. Thus a 1% increase in nonperforming loans causes a decrease of 5.7% in relative conditional bank efficiency. Therefore, this variable is important from an economic point of view to assess bank efficiency, in line with the recent literature results for other emerging markets. See Das and Ghosh (2006).

7. Summary and Conclusions

We introduced the notion of a probabilistic frontier and the companion definitions of input and output measures of efficiency. The concepts lead to FDH measures of efficiency and are a less restrictive alternative than the two-stage DEA to assess the statistical significance of contextual variables.

For a sample of 94 Brazilian banks we test whether bank type, specialization, size, type of ownership and nonperforming loans are relevant covariates to explain bank efficiency measured by FDH conditional measures of efficiency. We found that nonperforming loans are statistically significant and negatively correlated with bank efficiency. There is no evidence that size, specialization, and ownership affect the frontier. The notion of probability-based efficiency seems to be more informative than the standard DEA to detect potential performance failure in the Brazilian banking system. This disagreement with DEA suggests nonconvexity of the underlying technology or lack of separability relative to nonperforming loans. The statistical results also imply that either the bad management or the bad luck hypothesis, or both, cannot be rejected.

Nonperforming loans have been to a large extent a problem for the Brazilian banking system. The statistical significance of q in our study supports this assertion. The quantitative effect of nonperforming loans is measured by the elasticity of conditional bank efficiency. A relative increase of 1% in nonperforming loans implies, on average, a significant relative decrease of 5.7% in conditional bank efficiency.

References

Banker, R. D. (1993). Maximum likelihood consistency and DEA: A statistical foundation. *Management Science*, 39:1265–1273.

Banker, R. D. & Natarajan, R. (2004). Statistical tests based on DEA efficiency scores. In *Handbook of Data Envelopment Analysis*. Kluwer, New York.

Banker, R. D. & Natarajan, R. (2008). Evaluating contextual variables affecting productivity using data envelopment analysis. *Operations Research*, 56:48–58.

Berger, A. N. & DeYoung, R. (1997). Problem loans and cost efficiency in commercial banks. *Journal of Banking & Finance*, 21:849–870.

Berger, A. N. & Humphrey, D. B. (2000). Efficiency of financial institutions: International survey and directions for future research. In *Performance of Financial Institutions: Efficiency, Innovation, Regulation*. Cambridge, UK.

Bernstein, D. (1996). Asset quality and scale economies in banking. *Journal of Economics and Business*, 48:157–166.

Campos, M. B. (2002). Produtividade e eficiência do setor bancário privado brasileiro de 1994 a 1999. Master's thesis, EAESP-FGV, São Paulo.

Cazals, C., Florens, J. P., & Simar, L. (2002). Nonparametric frontier estimation: A robust approach. *Journal of Econometrics*, 106:1–25.

Coelli, T., Rao, D. S., O'Donnel, C. J., & Battese, G. E. (2005). *An Introduction to Efficiency and Productivity Analysis*. Springer, New York, second edition.

Colwell, R. J. & Davis, E. P. (1992). Output and productivity in banking. *Scandinavian Journal of Economics*, 94:111–129. Supplement.

Daraio, D. & Simar, L. (2005). Introducing environmental variables in nonparametric frontier models: A probabilistic approach. *Journal of Productivity Analysis*, 24:93–121.

Daraio, D. & Simar, L. (2007). *Advanced Robust and Nonparametric Methods in Efficiency Analysis*. Springer, New York.

Das, A. & Ghosh, S. (2006). Financial deregulation and efficiency: An empirical analysis of Indian banks during the post reform period. *Review of Financial Economics*, 15:193–221.

Hanes, M. & Thompson, S. (1999). The productivity effects of bank measures: Evidence from the UK building societies. *Journal of Banking and Finance*, 23:825–846.

Hoff, A. (2006). Second stage DEA: Comparison of approaches for modelling the DEA score. *European Journal of Operational Research*, 181:425–435.

Nakane, M. I. (1999). Productive efficiency in Brazilian banking sector. Discussion Paper 20/99, IPE-USP, São Paulo.

Nakane, M. I. & Weitraub, D. B. (2005). Bank privatization and productivity: Evidence for Brazil. *Journal of Banking and Finance*, 29:2259–2289.

Park, B., Simar, L., & Weiner, C. (2000). The FDH estimator for productivity efficiency scores: Asymptotic properties. *Econometric Theory*, 16:855–877.

Rossi, S., Schwaiger, M., & Winkler, G. (2005). Managerial behavior and cost/profit efficiency in the banking sectors of central and eastern European countries. Working Paper 96, Austrian National Bank.

SAS (2007). Statistical analysis system. V. 9.1.4, SAS-Insight.

Silva, T. L. & Jorge-Neto, P. M. (2002). Economia de escala e eficiência nos bancos brasileiros após o real. *Estudos Econômicos*, 32:577–620.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

Simar, L. & Wilson, P. (2007). Estimation and inference in two-stage, semi-parametric models of production process. *Journal of Econometrics*, 136:31–64.

Souza, G. S. (1998). *Introdução aos Modelos de Regressão Linear e Não-Linear*. Embrapa-SPI, Brasília.

Souza, G. S. & Staub, R. B. (2007). Two-stage inference using DEA efficiency measurements in univariate production models. *International Transactions of Operations Research*, 14:245–258.

Souza, G. S., Staub, R. B., & Tabak, B. M. (2006). Assessing the significance of factors effects in output-oriented DEA measures of efficiency: An application to Brazilian banks. *Revista Brasileira de Economia e de Empresas*, 6:7–20.

Wheelock, D. C. & Wilson, P. W. (2003). Robust nonparametric estimation of efficiency and technical change in U.S. commercial banking. Working Paper 037A, Federal Reserve Bank of St. Louis.

Williams, J. (2004). Determining management behavior in European banking. *Journal of Banking and Finance*, 28:2427–2460.