JCC
Journal of
CENTRUM
Cathedra ™

CENTRUM
CENTRO DE NEGOCIOS

Using Random Forests and Logistic Regression for Performance Prediction of Latin American ADRS and Banks

by

Dr. Germán Creamer*

Ph. D. in Economics, University of Notre Dame, USA
Ph. D. in Computer Science (Computational Finance), Columbia University, USA
Lecturer, Columbia University, NY, USA

Abstract

In the paper, random forests and logistic regressions' support of financial analysis functions' predictive tool to forecast corporate performance and rank accounting and corporate variables according to their impact on performance is demonstrated. Ten-fold cross-validation experiments are conducted on one sample each of Latin American depository receipts (ADRs) and Latin American banks. Random forests indicate that the most important variables that affect ADRs performance are size and the law-and-order tradition; the most important variables that affect banks are size, long-term assets to deposits, number of directors, and efficiency of the legal system. The interpretation of predictive models for a small sample improved when the capacity of random forests to rank and predict with the parameters of a logistic regression were combined.

Keywords: financial analysis, machine learning, random forests, logistic regression, data mining.

Using Random Forests and Logistic Regression for Performance Prediction of Latin American ADRS and Banks

Many of the recent bankruptcy scandals in publicly held US companies such as Enron and WorldCom are inextricably linked to a conflict of interest between shareholders (principals) and managers (agents). The conflict of interest, called the principal agent problem in finance literature, stems from tension between the interests of the investors in increasing the value of the company and the personal interests of the managers.

The principal agent problem is expected to have an important effect on company performance and efficiency. This is true of the financial markets of countries in the process of development, called emerging markets, because of lax security regulations. The study of corporate governance in emerging markets is especially important because these markets have become integrated into the major world financial centers. Emerging market stocks are represented in the US financial market through American Depository Receipts (ADRs). An ADR is a stock that represents a certain number of shares of a foreign company in the major US stock markets such as the NYSE. Only Latin American ADRs and banks domiciled in Latin American countries will be the focus of the paper, however.

In the paper, how random forests (Breiman, 2001b) and logistic regression are combined to evaluate corporate performance is demonstrated. A predictive model is created for evaluating whether a company's ADR performance or a bank's efficiency is above or below par as a function of the main corporate governance factors, such as those described above, and of selected accounting ratios that are known to be important for evaluating corporate governance risk. Random forests and logistic regression are used as the learning algorithms of the predictive model. In a previous paper, Creamer and Freund (2004) evaluated Latin American corporate performance using boosting.

Previous studies about US securities (Altman, 1968, 1974, 1989; Altman, Caouette, & Narayanan, 1998; Barr,

Seiford, & Thomas, 1994; Beaver, 1966; Collins & Green, 1982; Chen & Lee, 1995; Clarke & McDonald, 1992; Goudie & Meeks, 1992; Hudson, 1997; Lane, Looney, & Wansley, 1986; Hing-Ling, 1987; Moyer, 1977; Ohlson, 1980; Pinches & Mingo, 1973; Queen & Roll, 1987; Rose & Giroux, 1984; Zavgren, 1983) have used linear discriminant analysis or logistic regression for the evaluation of financial distress, bankruptcy, and bond and loan performance. This analysis is based on estimating the parameters of an underlying stochastic system that is usually assumed to be a linear system. A major limitation of the methodology is that non-linearity has to be incorporated manually.

In contrast, machine learning methods such as random forests avoid the question of modeling the underlying distribution and focus on making accurate predictions for some variables given other variables. Breiman (2001a) contrasted the two approaches respectively as the data modeling culture and the algorithmic modeling culture. According to Breiman, while most statisticians adhere to the data-modeling approach, people in other fields of science and engineering use algorithmic modeling to construct predictors with superior accuracy. For Breiman, the main drawback of algorithmic modeling is that generated representations are hard to interpret.

The objective of the paper is to demonstrate how financial analysis can be conducted using random forests as a predictive tool for corporate performance and logistic regression to improve the interpretability of the results. The analysis is conducted with emerging market companies, American Depository Receipts (ADRs), and banks domiciled in Latin American countries.

A predictive model is created for evaluating whether a company's performance or a bank's efficiency is above or below par as a function of the main corporate governance factors and selected accounting ratios known to be important in evaluating corporate performance. Random forests and logistic regression are used as the learning algorithms of the predictive model. The remainder of the paper is organized as follows: In the Learning Methods section, the predictive methods used are presented; in the Data and Variables section, the data and main variables are described; in the Experiments section, the experiments conducted are described in detail; in the Results section, the results of the forecast are presented; in the Financial Interpretations section, the results are discussed from a financial perspective; and in the Final Comments and Conclusions section, the conclusions are presented.

Learning Methods

In this section, the main learning methods used are introduced. The most important learning algorithms with respect to the research are random forests and logistic regression.

Random Forests

Random forests is a variant of bagging decision trees also proposed by Breiman (2001b), for which a free computer code is available. The algorithm was chosen because it presents the best publicly available combination of decision trees and bagging.

The algorithm generates multiple trees (θ_i) from the training data and from a random vector (x) sampled independently with the same distribution for any tree that is part of the forest. As a result, each tree generates a classifier $h(x,\theta_i)$. The majority votes of all the trees determine the predicted class. When the number of trees is very large, the generalization error for forests converges. Breiman (2001b) indicated that the accuracy of random forests is as good as Adaboost or better. Random forests generates a standardized score, z score that indicates the importance of each variable in the final classification.

Multiple Logistic Regression

The multiple logistic regression (Hastie, Tibishirani, & Friedman, 2003) models the posterior probabilities $Pr(Y = l|X_i)$ of L classes Y using linear regression in the observed values X_{ij} of the input variable $X_i = (X_{i1}...X_{in})$ of the feature j

$$Pr(Y = l|X_i) = \frac{e^{\sum_{j=1}^{n} \beta_{ij} X_{ij}}}{1+e^{\sum_{j=1}^{n} \beta_{ij} X_{ij}}}$$

where the odds ratio is $\frac{Pr(Y=l|X_i)}{Pr(Y=L|X_i)}$.

Le Cessie and Houwelingen (1992) proposed a method based on ridge estimators to improve parameters' estimates of the logistic regression model to avoid over fitting, reduce test errors, and generate interpretable parameters and odds ratios. According to Le Cessie and Houwelingen, the optimal ridge parameter minimizes the mean test error rate of several cross-validation tests of the model with different ridge parameters. The method was used to choose the optimal ridge parameter for the tests using the software Weka (Witten & Frank, 2005) as presented in Experiments section.

Data and Variables Description

Data

The data used in the experiments were from (a) Latin American ADRs (LAADR), and (b) Latin American banks (LABANKS).[1]

The first dataset is called LAADR because it is a sample of 51 stocks domiciled in Latin American countries, namely, Argentina, Brazil, Chile, Colombia, Peru, Mexico, and Venezuela, that have issued ADRs of level II and III for the year 1998. Level I ADR is least restricted with

respect to the required compliance with U.S. regulations, so level I ADR has not been included in the analysis. Level II ADRs correspond to foreign companies that list their shares on NASDAQ, AMEX, or NYSE. These companies must fully obey the registration requirements of the SEC, including compliance with US GAAP. Level III ADRs refer to foreign companies that issue new stocks directly in the United States. This means that they have the same compliance requirements as any U.S. public company and are therefore the most regulated. ADRs from countries on the list of emerging markets database (EMDB) of the International Finance Corporation (IFC) were chosen.[2]

The financial information from COMPUSTAT for the year 1998 was obtained. The information about the value of market capitalization is from CRSP, and is compared with information from the NYSE. Corporate governance information was extracted, such as lists of directors, executives, and major shareholders, from the proxy statements published at Disclosure, Edgar, and companies' websites for the year 1998. In the case of LAADR, *insider ownership* is defined as the ownership of a company by a CEO, managers and relatives of the CEO, and members of the board of directors.

The second dataset is called LABANKS because it is a list of 104 Latin American banks. LABANKS consists of banks headquartered in Argentina, Brazil, Chile, Colombia, Peru, Ecuador, and Bolivia, which represent about 80% of the total assets of the private sector in the major Latin American countries.[3] Corporate bank information was obtained from Internet Securities Inc., central banks, regulators, and company websites. Financial as well as corporate information, similar to that collected for ADRs, was collected. The sample of banks was restricted by the availability of corporate finance records. Most of the financial data was from 2000. A few companies that merged or disappeared in 1998 were included using the financial statements of 1997. Corporate information was gathered for the period 1998-2000. Considering that the information about ownership structure is relatively stable, major consistency problems were not apparent.

The period 1998-2000 was selected because the period was a transition period that preceded a financial crisis in Argentina and was during or after financial crises in Ecuador, Brazil, Venezuela, and Mexico. That the results of the analysis could have been affected by these financial crises is understood; however, it is difficult to find a period in the recent Latin American history where the major countries of the region were not going through a macroeconomic or financial crisis.

Variables Description

Independent Variables or Features: Corporate Governance Factors and Accounting Indexes

In the experiments described in the next sections, the following were independent variables or features of the machine learning algorithms: For corporate governance variables of ADRs and banks, the percentage of insider ownership (T Insider) was included because the separation of ownership and control was seen as an opportunity for managers to accumulate wealth at the expense of shareholders. The next group of variables included for LAADR and LABANKS were those related to the structure of the board of directors, namely, outsiders on the board of directors (PartOutBOD), natural logarithm of the size of the board of directors (LnDIR), and the double role of the CEO as chair of the board of directors and manager (ChairCEO). Among the variables, outsiders on the board of directors was preponderant. Fama (1980) and Fama and Jensen (1983) explained how the separation of control and security ownership can be an efficient structure because professional outside directors may limit the power of managers to expropriate the residual claimants' interests. The size of the board of directors is also a relevant variable, according to Yermack (1996) as well as Fuerst and Kang (2004), because it has an inverse association with firm value in the case of large U.S. industrial corporations. Lipton and Lorsch (1992) and Jensen (1993) recommended that companies limit board membership to no more than seven or eight members. In addition, Jensen suggested that companies should separate the roles of CEO and chair because of the need for independence. If the CEO is also chair of the board, the dual function may have a negative impact on performance. Even more, Jensen recommended including active investors who hold a large equity or debt position in a company and take part in strategic decisions. Institutional ownership (InstPart) was included as another variable because large institutional shareholders act as monitors of managers' actions. Results might be ambiguous if there is insider ownership or hidden investment because large shareholders may often manage the firm for their own interests and not for the benefit of the majority of small shareholders.

For LAADR and LABANKS, corporate governance indicators were also included at the country level (La-Porta, de Silanes, Shleifer, & Vishny, 1998) and included efficiency of the judicial system [EfficiencyJudicialSystem], rule of law [RuleOfLaw], risk of expropriation [RiskOfExpropriation], risk of contract repudiation [RiskOfContractRepudiation], corruption [Corruption], quality of accounting system [Accounting], and legal system [English/French]. Based on these indicators, La-Porta et al. found that common-law (English) countries have the strongest legal protection for investors and French-civil law countries have the weakest. These variables were included because separating the effect of country variables from company variables was an objective of the research.

Groups of accounting variables well known for their predictive power were selected for all companies and were indirect indicators of corporate governance variables. The accounting variables were the following: the logarithm of market capitalization (LnMarketCap) for ADRs and an equity index per country as a proxy for size for Latin American banks;[4] long-term assets to sales ratio (KS) for ADRs and long-term assets to deposits (KD) for banks

for their effect in the reduction of the agency conflict;[5] debt to total assets ratio (DebtRatio) as a capital structure indicator; operating expenses to sales ratio (Efficiency) as an efficiency or agency cost indicator;[6] operating income to sales ratio (YS) as a market power proxy and indicator of cash available from operations; and capital expenditures to long-term assets ratio (IK)[7] as a proxy for the relationship between growth and the possibility of investing in discretionary projects. A large IK ratio may indicate agency problems if managers are developing new projects that increase their power but do not add market value to the company. Region and sector were used as indicators of the geographical area and industrial sector in which the company operated (see Table 1).[8]

Table 1
Variables Used for Corporate Governance Experiments

Indicator	Definition	Type of Companies
TobinQ	Tobin's Q, which is the ratio of the market value to the replacement cost of assets. We use a proxy for Tobin's Q as the ratio of book value of debt plus market value of common	LAADR
PartOutBOD	% outsiders on the board of directors	LAADR, LABANKS
LnDIR	Natural logarithm of board size	LAADR, LABANKS
InstPart	% institutional ownership	LAADR, LABANKS
T_Insider	% insiders' ownership. In the case of LAADR and the Latin American banks, insider ownership is defined as ownership of a company by the CEO, managers, or relatives of the CEO, and members of the board of directors	LAADR, LABANKS
ChairmanCEO	1 if CEO is chairman, 0 otherwise	LAADR, LABANKS
LnMarketCap	Natural logarithm of market capitalization, used to measure firm size	LAADR
KS or KD	Ratio of long term assets (property, plant and equipment) to sales (KS) for LAADRs, and to deposits (KS) for LABANKS. This ratio is considered for its effect in the reduction of the agency conflict because these assets can be monitored very easily, and they can become collateral for the development of new projects	LAADR, LABANKS
YS	The ratio of operating income to sales	LAADR, LABANKS
DebtRatio	The ratio of debt to total assets, used as a capital structure variable. Emerging market are much less liquid than those of developed countries. Hence, firms may give more importance to debt, rather than equity, as a source of capital	LAADR, LABANKS
Equity index	Index of equity according to country of residence. This is a measure of size applied to LABANKS	LABANKS
Efficiency	The ratio of operating expenses to sales. This is the efficiency ratio and works as a proxy for market power. It also indicates cash flow available for management use. Similiarly, this efficiency ratio may also reveal agency costs or agency conflicts. (This is different from the DEA technical efficiency indicator)	LAADR, LABANKS
IK	The ratio of capital expenditures to long terms assets (stocks of property, plant and equipment)	LAADR, LABANKS
AvgParticipation	Measure of ownership concentration. This is calculated as the average of the participation of the three largest shareholders per firm	LAADR, LABANKS
English	If the firm is domiciled in a country whose legal regime is part of the common law or English law legal family according to La Porta et al. (1998)	LAADR, LABANKS
French	If the firm is domiciled in a country that is part of the Napoleonic or French legal family according to La Porta el al.	LAADR, LABANKS
RuleOfLaw	Law and order tradition according to the agency International Country Risk (ICR). Scores are from 0 to 10. Lower values indicate that a country is characterized by less tradition of law and order	LAADR, LABANKS
Corruption	Indicator of level of government corruption according to ICR. Low levels indicate higher corruption, such as solicitation of bribery by government officials	LAADR, LABANKS
EfficiencyJudicialSystem	Index about the level of efficiency of the legal system according to the agency Business International Corp. Scale is from 0 to 10. Lower values correspond to lower efficiency levels	LAADR, LABANKS
RiskofExpropriation	Risk of confiscation or nationalization according to ICR. Sales is from 0 to 10. Lower values imply higher risks	LAADR, LABANKS
RiskofContractRepudation	Risk of modification of a contract by economic, social or political reasons as defined by ICR. Lower values correspond to higher risks	LAADR, LABANKS
Accounting	Income statements, balance sheets, fund flow statement, accounting standards, stock data and special items. For each country, a minimum of three companies were studied	LAADR, LABANKS

Note. Third column indicates the type of company or dataset where each variable is used: LAADR for Latin American ADRs, and LABANKS for Latin American banks. Corporate governance variables at the country level are from La-Porta et al. (1998). These variables are English, French, RuleOfLaw, Corruption, EfficiencyJudicialSystem, RiskOfExpropriation, RiskOfContractRepudiation, and Accounting.

Measuring company performance

Tobin's Q performance measure of the value of intangibles or the real value created by management for ADRs was used. Tobin's Q is the ratio of the market value of assets to the replacement cost of assets.[9] High values of Tobin's Q indicate that more value has been added or there is an expectation of greater future cash flow. Hence, Tobin's Q captures the impact of management quality on performance. Any difference in Tobin's Q from 1 indicates that the market perceives the value of total assets to be different from the value required to replace physical assets. The value of internal organization, management quality, or expected agency costs is assumed to explain the difference. Values of Tobin's Q above 1 indicate that the market perceives the firm's internal organization as effective in manipulating company assets, while a Tobin's Q below 1 shows the market expects high agency costs. A proxy for Tobin's Q was used, namely, the ratio of book value of debt plus market value of common stocks and preferred stocks to total assets.[10]

Measuring efficiency of Latin American banks

For Latin American banks, an efficiency measure based on DEA instead of Tobin's Q was used because some of the banks under study are not public companies or do not participate in very illiquid markets. Efficiency indicators also calculate the agency costs to the firm. Conflicts between managers and shareholders may arise when operating costs increase in relation to a fixed output.

The present banking literature gives significant importance to the efficiency evaluation of financial institutions, applying parametric and nonparametric frontier analysis techniques to a specific company as part of an industry or firm's branches. Frontier analysis, based on optimization methodologies, selects the "best practice" firms or areas of a firm, obtains an efficiency score, and recognizes areas where there is overuse of inputs or underproduction of outputs within complex operations. Regulators use these techniques (Bauer, Berger, Ferrier, & Humphrey, 1998) to recognize the efficiency gain of a merger between two financial institutions. Frontier analysis can also be used to relate the level of risk that the firm is taking to its overall efficiency and establish "benchmarks" for financial institutions based on a best-practice frontier. The benchmarks can be established by regulators and managers who want to ensure that the firms they run are competitive nationally or internationally as compared with the rest of the industry (Berger & Humphrey, 1997).

From an economics point of view, the study of efficiency has been influenced by Leibenstein (1978) and his concept of X-efficiency. The economic concept of efficiency includes technical efficiency and implies efficiency in allocation, where the firm must choose an optimal combination of input and output that minimizes costs or maximizes output based on the production technology as well as relative market prices. X-efficiency refers to technical efficiency. Examples of the approach appear in the early nonparametric frontier models (Chames, Cooper, & Rhodes, 1978) and in some of the early parametric frontier models such as proposed by Aigner, Lovell, and Schmidt (1977).

The frontier approaches used to measure efficiency can be based on the following:

1. Nonparametric methods: Data envelopment analysis (DEA) is a linear programming technique to measure X-efficiency where the set of best-practice (frontier) observations are those for which no other combination of firms has as much of every output, given the firm's input, or as little of every input, given the firm's output. The firm's subject of study receives a score based on how efficient it is in relation to the best-practice firm. The method does not require a specific functional form between production outputs and inputs. However, the drawbacks to this method are that it assumes that there is not random error that leads to overestimating inefficiency, and the method suffers of the curse of dimensionality with several inputs and outputs or with relatively small samples.[11]

2. Parametric methods:
   a) The stochastic frontier approach (SFA), or the econometric frontier approach, imposes a functional form, such as the cost function, and recognizes the random error.
   b) The thick frontier approach (TFA) is similar to the SFA, but the estimations are based on the best performers in the data as estimators of the best-practice cost function for the whole group.
   c) The distribution free approach (DFA) handles a cost function, as do the two previous techniques, but assumes that there is an average efficiency and that the random error term tends to be eliminated.

Stochastic or parametric methods use prices and focus on economic efficiency. For this reason, Bauer et al. (1998) suggested that parametric methods should be used for regulatory purposes. These methods differ in the way they treat random errors and how they separate errors from inefficiency. The disadvantage of parametric methods is that the frontier must have a functional form such as the cost function.

Efficiency studies in the financial sector have been conducted mainly in the United States (Borger, Ferrier, & Kerstens, 1998; DeYoung, 1997; Mester, 1997; Thompson, Brinkmann, Dharmapala, Gonzalez-Lima, & Thrall, 1998), and on a smaller scale in Europe (Athanassopoulos, 1997; Grifell-Tatje & Lovell, 1997; Lozano, 1997), Canada (Schaffnit, Rosen, & Paradi, 1997), Saudi Ara-

bia (Al-Farj, Alidi, & Bu-Bshait, 1993), Tunisia (Chafai, 1997), Turkey (Zaire, 1995), and India (Bhattacharyya, Lovell, & Sahay, 1997).[12] In Latin America, efficiency studies in the banking sector have been scarce. Thompson et al. (1998), Pastor, Prez, & Quesada (1997), and Berger and Humphrey (1997) have compared international studies about banking efficiency. Previous studies that have addressed the relationship between efficiency and corporate governance structure in Latin America were not apparent.

DEA measures the performance of each producer relative to the best observed practice among *k* producers. The DEA frontier is a piecewise linear combination that connects the set of best-practice observations, creating a convex production possibilities set. The rest of the firms that are not in the frontier are ranked accordingly. DEA calculation implies the minimization of a weighted sum of inputs in relation to a weighted sum of outputs

$$\min_{u,v} \frac{v_T x_0}{u^T y_0}$$

subject to

$$\frac{v_T x_i}{u^T y_i} \geq 1$$
$$u, v \geq 0$$

where: *i*= l,...,0,...k

(x_0, y_0): input-output vector of the firm that is evaluated

(x_i, y_i): input-output vector of *i*th firm in the sample

u: vector of weights given to output

v: vector of weights given to input

This minimization problem can also be expressed as a linear programming problem

$$\min_{u,v} v_T x_0$$

$$u^T y_0 = 1$$
$$v_T x_i \geq u^T y_i \quad \text{where i= 1,...,0,...k}$$
$$u, v \geq 0$$

and then as the dual linear programming "envelopment" problem

$$\max_{\theta,\gamma} \theta$$

subject to

$$X\gamma \leq x_o$$
$$\theta y_o \leq Y\gamma$$
$$\gamma \geq 0$$

X is an *n* by *k* input matrix, *Y* is an *m* by *k* output matrix, γ is a *k* by 1 intensity vector, and X_i and Y_i are the columns of the input and output matrix respectively. θ is a radial measure of technical efficiency. An optimal firm will have its efficiency measure (θ) equal to 1. If it is more than 1, it can still increase its output with the same unit of input. This version of DEA is output oriented, assumes constant returns to scale and was proposed by Chames et al. (1978; see also Lovell, 1993).

Efficiency for the Latin American banking sector was calculated using the DEA with different variations, such as input-oriented or output-oriented; constant, nondecreasing, or variable returns to scale; and stochastic DEA. Parametric methods would be more appropriate if a consistent dataset of prices, expenditures, and income across different countries and banks were available. Because most of the Latin American banks under study are not public companies and have different accounting standards, the DEA approach was used. The DEA approach focuses on technology efficiency, minimizing the potential distortion generated by inconsistent prices and standards across different countries.

Only output-oriented constant returns to scale were used as measures of banking efficiency, and results were consistent with the results obtained with the other methods. For calculations, the main accounting items of the income statement were selected because they were the only indicators that are available for the banks under study. As input, interest-paying deposits and non-interest expenses were used, which included personnel, administrative, commissions, and other non-interest operating costs. As output, total income was used, which included interest and non-interest income. Banks were ranked according to this measure, country by country. If a bank showed a great level of inefficiency, a potential agency conflict may have been present. The definition of the inputs used in the efficiency calculations was standard for most of the efficiency studies for banks. The output measure was also typical in the literature. Other studies have used different outputs (e.g. installment loans, real estate loans, commercial and industrial loans, demand deposits, and off-balance sheet activities). Variations in the efficiency method, inputs, outputs, or periods of analysis may lead to different results. Even though the quality and availability of data to the methods used may have been restricted, findings of several authors obtained similar efficiency results using different parametric methods (Bauer, Berger, & Humphrey, 1993; Berger & Hannan, 1998; Berger & Mester 1997; Hasan & Hunter, 1996) or comparing parametric and nonparametric methods (Eisenbeis, Ferrier, & Kwan, 1997; Ferrier & Lovell, 1990; Resti, 1997).

Experiments

Random forests and logistic regression were used (see Random Forests section) to classify stocks above and below the median of Tobin's Q for LAADR companies and the DEA technical efficiency indicator for LABANKS (see Measuring Efficiency of Latin American Banks section). As independent variables, the accounting and corporate governance variables introduced were used (see Independent Variables or Features section). In the LAADR

sample, the median of the Tobin's Q was very close to 1. The results can be interpreted as a classification between stocks with a market value of assets above (Tobin's Q > 1) or below (Tobin's Q < 1) costs of replacement. For LABANKS, the classification was between more efficient and less efficient banks. The efficiency indicators for each country were calculated because of the differences between their accounting systems. The banks' efficiency was calculated in relation to peers in their respective countries.

The results of ADTs are interpreted as companies with positive scores and high Tobin's Q, and banks that are efficient institutions; or companies with negative scores and low Tobin's Q and inefficient banks.

Variables that indicated multi-collinearity were eliminated using the variance inflation factor (VIF). In general, variables with large VIF (>10) were removed. For LABANKS the eliminated variables were risk of contract repudiation, the legal system, the region, corruption, and the debt ratio. For LAADR, risk of expropriation, risk of contract repudiation, and the region were eliminated.

Random forests do not require cross-validation on held-out experiments to achieve an estimate of the test error. The test error is obtained with a regular run of random forests because each tree is generated using a bootstrap sample of the data. Only two-thirds of the bootstrap sample was used to generate each tree; the rest were left out for the test set.

Tests of random forests and logistic regression were conducted using the software Weka (Witten & Frank, 2005) and Random Forests V5.0 respectively.[13] Random forests experiments with 1,000 trees, and with four variables randomly selected at each node in order to reduce the test error for LAADR and LABANKS companies were conducted.

The logistic regression analysis (see Random Forests and Multiple Logistic Regression section) used Tobin's Q as the dependent variable for LAADR companies and the DEA technical efficiency indicator as the dependent variable for LABANKS (see Measuring Company Performance section). Besides the independent variables introduced (see Independent Variables or Features section), the multiple logistic regression also included indicator variables for industrial sectors. The multiple logistic regressions with ridge estimators were conducted to avoid overfitting and obtain interpretable parameters and odds ratios.

Following Le Cessie and Houwelingen (1992), the ridge parameters that minimized the mean test error rates or converged the error rates for ten-fold cross-validation tests were selected. Based on the criterion, the ridge parameters for LAADR and LABANKS were zero, so the original logistic regression was not changed (see Figures 1 and 2).

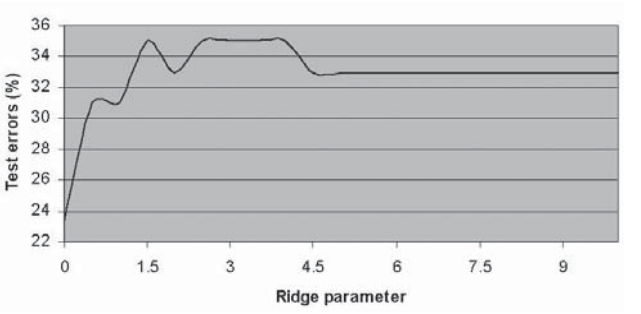

Figure 1. Cross-validated estimate of the mean of test errors for multiple logistic regression as a function of ridge parameters for LAADR.

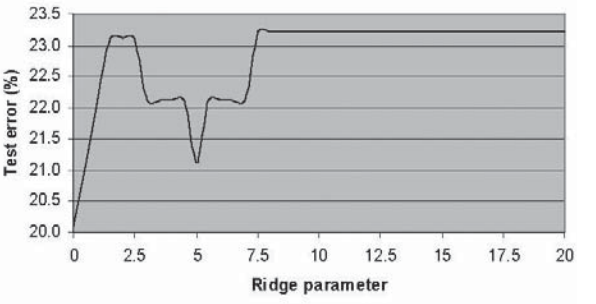

Figure 2. Cross-validated estimate of the mean of test errors for multiple logistic regression as function of ridge parameters for LABANKS.

Results

A significant difference among the learning algorithms for the case of LAADR and LABANKS did not exist (see Table 2).[14]

Table 2
Test Errors and Standard Deviations of Learning Algorithms When All Variables Are Included

	LAADR		LABANKS	
	Test error	St. dev.	Test error	St. dev.
Random forests	32.0%	16.9%	16.7%	17.6%
Logistic regression	23.3%	20.2%	20.1%	13.2%
Number observations	51		104	

One of the main restrictions of the research was the small sample used for the analysis because the dataset was restricted to Latin American companies with publicly available accounting and corporate governance indicators. The limited number of observations may have affected the predictive capacity of the learning algorithms. In addition, the small sample may explain why the ridge estimators could not improve the logistic regression model. Hence, the very large odds ratios observed in the logistic

regression may be appropriate for the model developed. However, the odds ratios are difficult to interpret. This limitation is partially compensated for by the capacity of random forests to rank the most important variables using a *z* score. The shortcoming of random forests is its interpretability because it generates many random trees to formulate its prediction. The interpretation of the results was completed using the ranking of random forests with the parameters of the logistic regression.

Table 3 indicates the importance of each variable according to random forests and logistic regression. For the LAADR dataset, the relevant variables are market capitalization, law and order tradition, capital expenditure to long-term assets ratio, operating expenses to sales ratio, and debt to total assets ratio. The relevant variables for the LABANKS dataset are equity index, long-term assets to deposits ratio, number of directors, efficiency of the legal system, and risk of expropriation. The odds ratios of logistic regression also confirm for LABANKS the relevance of random forests of the above variables, with the exception of size.

Table 3
Results for ADR and LABANKS

	LAADR								LABanks							
	Statistics				Logit		RF		Statistics				Logit		RF	
	Q25	Median	Q75	Mean	Coeff.	Odds ratios	z-score	Rank	Q25	Median	Q75	Mean	Coeff.	Odds ratios	z-score	Rank
LnMarketCap (Nat. log market capitalization)	5.44	6.73	7.49	6.57	-15.14	0.00	26	1	(Not used)							
Equity index	(Not used)								0.04	0.16	0.50	0.30	-7.64	0.00	35	1
IK (Capital expenditures/ long-term assets)	0.05	0.08	0.13	0.10			6	3	(Not used)							
Efficiency (Operating expenses / sales)	0.10	0.16	0.23	0.16			5	4	(Not used)							
YS (Operating income / sales)	0.13	0.23	0.35	0.25	-29.06	0.00			(Not used)							
Debtratio(Debt / total assets)	0.46	0.59	0.80	0.61			2	5	0.89	0.92	3.23	77.96				
KS (L.T. assets/sales)	0.73	1.44	2.20	1.82	10.88	53156.67										
KD (LT ass./deposits)									0.04	0.06	0.10	0.11	28	2.08E+12	33	2
TobinQ (Tobin's Q: performance)	0.91	1.04	1.44	1.38					(Not used)							
EfficiencyJudicialSystem (Effic. legal system)	6.00	6.00	7.25	6.50					6.00	6.25	6.75	6.43	-0.79	0.46	12	4
RuleOfLaw (Law and order tradition)	5.35	5.35	7.02	5.82	-3.63	0.03	6	2	2.50	6.32	6.67	5.27	-0.46	0.63	8	6
Corruption (Level of government corruption)	4.77	5.30	5.30	5.21	11.32	82222.46			5.00	5.18	6.02	5.45				
RiskOfExpropiation (Risk confiscation)	6.95	7.29	7.50	7.09					5.91	6.57	7.50	6.67	-0.75	0.47	10	5
RiskOfContractRepudiation (Contract change)	6.30	6.55	6.80	6.33					4.91	5.18	6.30	5.68				
PartOutBOD (% outsiders as directors)	60.0%	77.0%	87.0%	68.1%	-35.63	0.00	2	6	75.0%	94.4%	100.0%	84.8%	0.93	2.53	3	9
Avg Participation	(Not used)								0.50	0.93	1.00	0.75	-1.85	0.16	4	8
LnDir (Natural log number directors)	1.95	2.20	2.30	2.08	25.4	1E+11			1.79	2.20	2.40	2.10	-0.90	0.41	12	3
InstPart (% institutional equity ownership)	15.0%	44.0%	71.0%	43.2%	3.32	27.76			(Not used)							
T_Insiders (% insider's equity ownership)	0.0%	0.0%	2.0%	10.4%			0.02	7	0.0%	0.0%	1.2%	8.8%	-0.64	0.53	4	7

Note. This table reports statistics and results of predicting Tobin´s Q for LAADR and efficiency for LABANKS using logistic regression, and random forests. Country corporate governance variables are from La Porta et. al. (1998). RF: Random forests. z-score for random forests (Breiman 2001a) is the raw importance score divide by standard deviation. Q25: 25th. percentile. Logistic regression includes indicator variables to control for sector, although they are not included in the table. Variables that do not show any relevance are not included such as legal system, accounting, number of insiders in board of directors, and chairman as CEO. Corporate governance variables are in gray.

Financial Interpretation

Comparing the main variables selected for LAADR and LABANKS (see Table 3), the main distinctive variable is the size of the company measured by the logarithm of market capitalization for LAADR and the equity index for LABANKS. Large companies in emerging markets are likely to be oligopolies or monopolies in their areas of activity.

The accounting indicators, capital expenditures to long-term assets ratio, operating expenses to sales ratio, and debt to total assets ratio play a central role in the case of LAADR, and long-term assets to deposits ratio in the case of LABANKS. These indicators are important for revealing agency problems. The long-term assets are easy to monitor and can become collateral to finance new projects. However, if the level of long-term assets is too high, it may indicate inefficiency and overspending. An excessive amount of capital expenditures may allow managers to spend reserves on projects that benefit them directly instead of increasing the value of their companies. A high debt ratio for LAADR implies additional monitoring by creditors. Harvey, Linsc, and Roper (2004) found that in emerging market companies with significant agency conflicts, shareholders benefit from very close supervision of their debt. A large operating expense to sales ratio (efficiency ratio) may also indicate agency conflict.

Corporate governance variables seem to be more important for LABANKS than for LAADR. Government entities supervise and regulate banks intensely. In addition, the information published in the proxy statements of ADRs are not under the same strict control that financial statements are. As a result, it is possible that many ADRs

did not include relevant information about managers, ownership structure, and board composition due to the need to protect shareholders against potential kidnapping or assault. Only the major shareholders are registered.

The participation of outside directors is the second most important variable for LAADR. The finance literature indicated that outside directors supervise managers (Weisbach, 1988; Shivdasani 1993). Weisbach (1988) found that outsider-dominated boards were more likely to remove CEOs than firms with insider-dominated boards, especially when firms showed poor performance.[15] Denis and Sarin (1999) found that companies that increased the proportion of outsiders on the board of directors or reduced ownership concentration had above average returns in the previous year. However, Yermack (1996), MacAvoy, Cantor, Dana, and Peck (1983), Hermalin and Weisbach (1991), and Bhagat and Black (1999, 2002) found little correlation between the composition of boards of directors and performance. One possible explanation for the results is that the CEO hires outside directors; hence, directors do not dissent (Crystal 1991). Core, Holthausen, and Larcker (1999), who found that CEO compensation was a decreasing function of the share of inside directors and an increasing function of the share of outside directors chosen by the CEO, reinforce the hypothesis.

Inside directors also play an important role in the board of directors for making strategic planning decisions, reviewing functional performance by areas and, in some cases, evaluating important differences between CEOs' perspectives and firms' daily business.[16] Baysinger and Butler (1985) proposed that an optimal board of directors should have a combination of inside, independent, and affiliated directors. Bhagat and Black (2002) suggested that boards should not only be composed of independent directors because their findings showed that board independence does not improve performance and inside directors may bring the additional benefits delineated above.

In LAADR and LABANKS, insiders' equity ownership is also a relevant variable that affects performance. The logistic regression shows that the variable is negatively associated with efficiency in the case of LABANKS. Management with a high level of ownership is likely to steer corporate decisions toward its own interests at the expense of corporate interests. This could be true in the case of strong family groups that control a company. Family groups may use their great bargaining power to make corporate decisions that benefit companies in which they have a great interest. For example, banks may direct an important part of their loan portfolios to companies where managers or insiders have significant interests. If the investment is successful, managers benefit. Otherwise, government and depositors assume the loss, as occurred in the financial crisis of the Andean countries during the 1990s. Jensen and Meckling (1976), in their classic work, described how large investors as equity holders will benefit when the firm takes an excessive risk because of the potential benefit on the upside, while other stakeholders, such as the creditors, bear all the risk. Hence, a strong incentive exists to be a large shareholder in developing countries. Hermalin and Weisbach (1991) already proposed that agency costs increase with ownership, such as in the case of family firms. La-Porta, de Silanes, and Shleifer (1999) described the agency problem in these companies, which is that dominant family owner-manager may expropriate minority shareholders. However, expropriation is expensive; the cost of expropriation might be bigger than its potential benefit in the case of controlling shareholders, which explains why La-Porta, de Silanes, Shleifer, and Vishny (2002) found that firms with higher cash-flow ownership by the controlling shareholder and companies with better shareholder protection (common law countries) had higher valuation measured by Tobin's Q.

The efficiency of the legal system, the level of government corruption, and the law and order tradition are negatively associated with efficiency for LABANKS. The law and order tradition also has a negative relationship with performance for the LAADR. Large Latin American companies probably perform better in environments with a weak legal structure, with less of a tradition of law and order, and with higher levels of corruption because of the close family relationships that help influence government decisions in their favor. In the sample, more than two thirds of the high performing companies operated in countries with less of a tradition of law and order, while only a third smaller companies operated in under such conditions. The benefits of government-private sector connections may be less important for companies of a small size.

In the case of LAADR and LABANKS, the limited impact of the size of the board of directors and the composition of the board (percent of outsiders) on performance and efficiency give findings similar to what previous studies have indicated for the USA. Bhagat and Black (2002) did not find that board independence led to improved profitability after controlling for firm size, board size, industry effects, CEO stock ownership, ownership by outsiders, and size and number of outside, 5% blockholders.

Final Comments and Conclusions

Because of the emerging market's financial crisis of the late 1990s, and the recent corporate scandals in the United States, several multilateral institutions, such as the World Bank, are supporting transformation in the corporate governance structure of emerging markets. The resistance to structural reforms seems similar to the resistance that local business groups have shown when an economy is open. Local business groups have been able to capture an important rent thanks to the protection of an inward-oriented industrialization process. The current stage of institutional reform in emerging markets must consider the interest of either the small investor who

needs to protect her property rights or the bank's client who needs some capital to keep alive and would benefit from an adequate management of financial costs by her bank. A new corporate governance structure should protect the property rights of investors and producers through a regulatory framework that reorients the power of strong family groups and management to corporate endeavors that create value for the shareholders, and therefore, for society as a whole.

The research shows that random forests combined with logistic regression can facilitate the financial analysis functions as a predictive tool to forecast corporate performance and to rank accounting and corporate variables according to their impact on performance. The use of machine learning methods in finance requires time-series or cross-sectional data in order to calculate meaningful results. Indicators that do not have enough information cannot be quantified. As the research showed, random forests also worked adequately with small datasets. However, the test error and its variance increased as the size of the dataset decreased. It is suggested that companies that use random forests as an interpretative tool use large datasets, such as industrial surveys or compensation surveys and build their own internal dataset using the company's historical information.

Comparative regional studies have a major problem in terms of how to integrate data coming from different sources with different standards. The problem was implicit in the LABANKS dataset. Research about emerging markets can be improved by expanding the dataset and running the learning algorithms in subsets aggregated by regions or corporate governance systems.

References

Aigner, D., Lovell, C., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics, 6,* 21-37.

Al-Farj, T. Alidi, A., & Bu-Bshait. K. (1993). Evaluation of bank branches by means of data envelopment analysis. *International Journal of Operations and Production Management, 13*(9), 45-52.

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance, 23*(4), 589-609.

Altman, E. I. (1974). Evaluation of a company as a going concern. *Journal of Accountancy, 138*(6), 50-57.

Altman, E. I. (1989). Measuring corporate bond mortality and performance. *Journal of Finance, 44*(4), 909-922.

Altman, E. I., Caouette, J. B., & Narayanan, P. (1998). Credit-risk measurement and management: The ironic challenge in the next decade. *Financial Analysts Journal, 54*(1), 7-11.

Athanassopoulos, A. (1997). Service quality and operating efficiency synergies for management control in the provision of financial services: Evidence from Greek bank branches. *European Journal of Operational Research, 98*(2), 300-313.

Barr, R., Seiford, L. M., & Thomas, F. (1994). Forecasting bank failure: A non-parametric frontier estimation approach. *Recherches Economiques de Louvain, 60*(4), 417-429.

Bauer, P., Berger, A., & Humphrey, D. (1993). *Measurement of productive efficiency: Techniques and applications.* New York: Oxford University Press.

Bauer, A., Berger, P., Ferrier, G., & Humphrey, D. (1998). Consistency conditions for regulatory analysis of financial institutions: A comparison of frontier efficiency methods. *Journal of Economics and Business, 50*(2), 85-114.

Baysinger, B., & Butler, H. (1985). Corporate governance and the board of directors: performance effects of changes in board composition. *Journal of Law, Economics and Organization, 1,*101-204.

Beaver, W. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research, 4,* 71-111.

Berger, A., & Hannan, T. (1998). The efficiency cost of market power in the banking industry: A test of the quiet life and related hypotheses. *Review of Economics and Statistics, 81,* 454-465.

Berger, A., & Humphrey, D. (1997). Efficiency of financial institutions: International survey and directions for future research. *European Journal of Operational Research, 98*(2), 175-212.

Berger, A., & Mester, L. (1997). Beyond the black box: What explains differences in the efficiencies of financial institutions? *Journal of Banking and Finance, 21,* 895-947.

Bhagat, S., & Black, B. (1999). The uncertain relationship between board composition and firm performance. *Business Lawyer, 54,* 921-963.

Bhagat, S., & Black, B. (2002). The non-correlation between board independence and long-term firm performance. *Journal of Corporation Law, 27,* 231-273.

Bhattacharyya, A., Lovell, C., & Sahay, P. (1997). The impact of liberalization on the productive efficiency of Indian commercial banks. *European Journal of Operational Research, 98*(2), 332-345.

Borger, B. D., Ferrier, G. D., & Kerstens, K. (1998). The choice of a technical efficiency measure on the free disposal Hull reference technology: a comparison using US banking data. *European Journal of Operational Research, 105*(3), 427-446.

Breiman, L. (2001a). Statistical modeling: The two cultures. *Statistical Science, 16,* 199-231.

Breiman, L. (2001b). Random forests. *Machine Learning, 45*(1), 5-32.

Chafai, M. (1997). Estimating input-specific technical inefficiency: The case of the Tunisian banking industry. *European Journal of Operational Research, 98*(2), 314-331.

Chames, A., Cooper, W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research, 2,* 429-44.

Chen, K., & Lee, J. (1995). Accounting measures of business performance and Tobin's Q theory. *Journal of Accounting, Auditing and Finance, 10*(3), 587-609.

Chung, K., & Pruitt, S. W. (1994). A simple approximation to Tobin's Q. *Financial Management, 23*(3), 70-74.

Clarke, D. G., & McDonald, J. B. (1992). Generalized bankruptcy models applied to predicting consumer credit behavior. *Journal of Economics and Business, 44*(l), 47-62.

Collins, R., & Green, R. D. (1982). Statistical methods for bankruptcy forecasting. *Journal of Economics and Business, 34,* 349-354.

Core, J. E., Holthausen, R. W., & Larcker, D. F. (1999). Corporate governance, chief executive officer compensation, and firm performance. *Journal of Financial Economics, 51*(3), 371-406.

Creamer, G. & Freund, Y. (2004). Predicting performance and quantifiying corporate governance risk for Latin American ADRs and banks. *Proceedings of the Financial Engineering and Applications Conference* (pp. 91-101). Cambridge, UK: Acta Press.

Creamer, G., & Freund, Y. (2005). *Using Adaboost for an equity investment/board balanced scorecard.* Unpublished paper presented at the Machine Learning in Finance Workshop in NIPS, Whistler, BC.

Crystal, G. (1991). *In search of excess: The overcompensation of American executives.* New York: W. W. Norton.

Denis, D., & Sarin, A. (1999). Ownership and board structures in publicly traded corporations. *Journal of Financial Economics, 52,*187-223.

DeYoung, R. (1997). A diagnostic test for the distribution-free efficiency estimator: An example using U.S. commercial bank data. *European Journal of Operational Research, 98*(2), 243-249.

Eisenbeis, R., Ferrier, G., & Kwan, S. (1997). *The informativeness of linear programming and econometric efficiency scores: An analysis using us banking data.* Working Paper, Bureau of Business and Economic Research. University of Arkansas, Fayetteville, AR.

Fama, E. (1980). Agency problems and the theory of the firm. *Journal of Political Economy, 88,* 288-307.

Fama, E., & Jensen, M. (1983). Separation of ownership and control. *Journal of Law and Economics, 26,* 301-325.

Ferrier, G. D., & Lovell, C. (1990). Measuring cost efficiency in banking: Econometric and linear programming evidence. *Journal of Econometrics, 46,* 229-245.

Fuerst, O., & Kang, S. (2004). Corporate governance, expected operating performance and pricing. *Corporate Ownership and Control, 1*(2): 13-30.

Goudie, A., & Meeks, G. (1992) *Forecasting and explaining corporate failure: A study of contingent financial statements* (Tech. Rep. N° AF1) Cambridge, UK: Cambridge University, Department of Economics.

Grifell-Tatje, E., & Lovell, C. (1997). The sources of productivity change in Spanish banking. *European Journal of Operational Research, 98*(2), 364-380.

Hall, M. J. (Ed.) (2001). *The regulation and supervision of banks: Regulation and efficiency in banking,* (Vol. 4). Cheltenham, UK: Edward Elgar.

Harvey, C. R., Linsc, K. V., & Roper, A. H. (2004). The effect of capital structure when expected agency costs are extreme. *Journal of Financial Economics, 74*(1), 3-30.

Hasan, I., & Hunter, W. (1996) Efficiency of Japanese Multinational Banks in the United States. *Research in Finance,* 14, 157-73.

Hastie, T., Tibishirani, R., & Friedman, J. (2003). *The elements of statistical learning.* New York: Springer.

Hermalin, B., & Weisbach, M. (1991). The effects of board composition and direct incentives on firm performance. *Financial Management, 20,* 101-112.

Hing-Ling, A. L. (1987). A five-state financial distress prediction model. *Journal of Accounting Research, 25,* 127-138.

Hudson, J. (1997). Company bankruptcies and births matter. *Applied Economics, 29,* 647-654.

Jensen, M. (1993). The modern industrial revolution, exit, and the failure of internal control systems. *Journal of Finance, 48,* 831-880.

Jensen, M., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics, 3,* 305-360.

Klein, A. (1998). Firm performance and board committee structure. *Journal of Law and Economics, 41,* 275-303.

La-Porta, R., de Silanes, F. L., & Shleifer, A. (1999). Corporate ownership around the world. *Journal of Finance, 54,* 471-517.

La-Porta, R., de Silanes, F. L., Shleifer, A., & Vishny, R. (1998). Law and finance. *Journal of Political Economy, 196:* 1113-1155.

La-Porta, R., de Silanes, F. L., Shleifer, A., & Vishny, R. (2002). Investor protection and corporate valuation. *Journal of Finance, 57,* 1147-1170.

Lane, W., Looney, S. W., & Wansley, J. W. (1986). An application of the Cox proportional hazards model to bank failure. *Journal of Banking and Finance, 10,* 511-531.

Le Cessie, S., & v. Houwelingen, J. C. (1992). Ridge estimators in logistic regression. *Applied Statistics, 41*(1), 191-201.

Leibenstein, H. (1978). *General X—Efficiency theory and economic development.* New York: Oxford University Press.

Lindenberg, E., & Ross, S. A. (1981). Tobin's Q ratio and industrial organization. *Journal of Business, 54,* 1-32.

Lipton, M., & Lorsch, J. (1992). A modest proposal for improved corporate governance. *Business Lawyer, 48,* 59-77.

Lovell, C. (1993). Production frontiers and productive efficiency. In H.O. Fried, C.A.K. Lovell and S.S. Schmidt (Eds.), *The Measurement of Productive Efficiency* (pp. 3-67). New York: Oxford University Press.

Lozano, A. (1997). Profit efficiency for Spanish savings banks. *European Journal of Operational Research, 98*(2), 381-394.

MacAvoy, P., Cantor, S., Dana, J., & Peck, S. (1983). ALI proposals for increased control of the corporations by the board of directors: An economics analysis. In *Statement of the Business Roundtable on the American Law Institute's Proposed Principles of Corporate Governance and Structure: Restatement and Recommendations.* New York.

Mester, L. (1997). Measuring efficiency at U.S. banks. European *Journal of Operational Research, 98*(2), 230-242.

Moyer, C. R. (1977). Forecasting financial failure: A re-examination. *Financial Management, 6*(1), 11-16.

Ohlson, J. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research, 18,* 1.

Pastor, J. M., Prez, F., & Quesada, J. (1997). Efficiency analysis in banking firms: An international comparison. *European Journal of Operational Research, 98*(2), 395-407.

Perfect S. B., & Wiles, K. W. (1994). Alternative construction of Tobin's Q: An empirical comparison. *Journal of Empirical Finance, 1,* 313-341.

Peterson P., & Peterson, D. (1996). *Company performance and measures of value added.* Charlottesville, VA: The Research Foundation of the Institute of Chartered Financial Analysts.

Pinches, G., & Mingo, K. (1973). A multivariate analysis of industrial bond ratings. *Journal of Finance, 28,* 1.

Queen, M., & Roll, R. (1987). Firm mortality: Using market indicators to predict survival. *Financial Analysts Journal, 43*(3), 9-26.

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, 20,* 221-50.

Rose, P. S., & Giroux, G. A. (1984). Predicting corporate bankruptcy: An analytical and empirical evaluation. *Review of Business and Economics Research, 19*(2), 1-12.

Rosenstein, S., & Wyatt, J. (1990). Outside directors, board independence, and shareholder wealth. *Journal of Financial Economics, 26,* 175-191.

Schaffnit, C., Rosen, D., & Paradi, J. C. (1997). Best practice analysis of bank branches: An application of DEA in a large Canadian bank. *European Journal of Operational Research, 98*(2), 269-289.

Shivdasani, A. (1993). Board composition, ownership structure and hostile takeovers. *Journal of Accounting and Economics, 16,* 167-98.

Thompson, R., Brinkmann, E. J., Dharmapala, P., Gonzalez-Lima, M., & Thrall, R. M. (1998). DEA/AR profit ratios and sensitivity of 100 large U.S. banks. *European Journal of Operational Research, 105*(3), 213-229.

Volpin, P. (2002). Governance with poor investor protection: Evidence from top executive turnover in Italy. *Journal of Financial Economics, 64,* 61-90.

Weisbach, M. (1988). Outside directors and CEO turnover. *Journal of Financial Economics, 20,* 431-460.

Wheelock, D. C., & Wilson, P. W. (2003). *Robust nonparametric estimation of efficiency and technical change in US commercial banking.* (Working paper 2003-037A). The Federal Reserve Bank of St. Louis.

Witten, I. H., & Frank, E. (2005). *Data mining: Practical machine learning tools and techniques* (2nd ed.). San Francisco: Morgan Kaufmann.

Yermack, D. (1996). Higher market valuation of companies with a small board of directors. *Journal of Financial Economics, 40,* 185-211.

Zaire, O. The effect of financial liberalization on the efficiency of Turkish commercial banks. *Applied Financial Economics, 5*(4), 257-264.

Zavgren, C. (1983). The prediction of corporate failure: the state of the art. *Journal of Accounting Literature, 2*(1), 1-37.

Footnotes

1 The data can be obtained from http://publications.creamer-co.com-a.googlepages.com/la_adrsbanks.

2 Standard and Poor acquired this database in January 2000, and it became the Standard and Poor's EMDB.

3 Venezuela's banks were not able to be included because the President of the Venezuelan Banking Association declined to supply any information to the research team, and asked member banks not to supply any corporate information to the research team due to the increased risk of kidnapping that its members would be subject to if the information were distributed.

4 The equity index was used instead of equity value because efficiency is calculated country by country. The effect of the relative size by country on efficiency instead of its absolute value is the focus.

5 Assets can be monitored very easily and can become collateral for the development of new projects or used to finance new acquisitions.

6 If operating costs are too high in relation to industry peers or previous years, it might be due to excessive perquisite consumption or other direct agency costs.

7 Operating expenses to sales ratio and operating income to sales ratio are calculated only for ADRs because these ratios are highly correlated with the efficiency indicator calculated for the banking sector. The capital expenditures to long-term assets ratio is also calculated only for the ADRs.

8 Sectors of activity for ADRs are defined according to the North American Industrial Classification System (NAICS).

9 The intangibles can also refer to other factors such as intellectual capital or the value of information technology. In the research, differences among countries and economic sectors where companies had similar technologies were controlled for. It was assumed that Tobin's Q reflects management quality. The discrimination between the contribution to the performance of top management and other intangible assets such as intellectual capital requires a more detailed analysis.

10 Several authors (Chung & Pruitt, 1994; Perfect & Wiles, 1994; Peterson & Peterson, 1996) indicated that the proxy is empirically close to the well known Lindenberg and Ross (1981) proxy. For international stocks, the information to calculate the Lindenberg and Ross proxy is very limited.

11 Wheelock and Wilson (2003) overcame many of the problems of DEA using an order m frontier where each bank is compared only with a reduced number of its peers.

12 See Hall (2001) for a collection of articles on bank efficiency from 1973 until 1998 for many countries.

13 A working version of Random Forests V5.0 can be obtained from http://stat-www.berkeley.edu/users/breiman/RandomForests.

14 Creamer and Freund (2005) extended the research to the S&P 500 companies in order to evaluate similar experiments with a larger dataset (2,278 observations).

15 In the case of Italy, the situation was different. Volpin (2002)

found that the probability of turnover and its relationship to performance was lower for executives who were part of the family of the controlling shareholder. Rosenstein and Wyatt (1990) found that announcements of outside directors were related to positive excess returns.

16 Klein (1998) found that inside director participation in investment committees correlated with better firm performance.

Author Note

Thanks go to Paul Spindt, Tom Noe, Kenneth Jameson, Greg Buchholz, John M. Trapani, Yoav Freund, Jorge Guillén, an anonymous referee and the participants in the 2003 Latin American Studies Association meeting. Thanks also to the institutional support of the Board of Regents of the State of Louisiana, and the research support of Carolina Gomez, Monica Garcia, Leonardo Serrano, Marcelo Gonzalez, Juan Carlos Otalora, and Julian Benavides. The opinions presented are the exclusive responsibility of the author.

* Correspondence concerning this article should be directed to Germán Creamer at gcreamer@pucp.edu.pe.