PREDICTING PERFORMANCE AND QUANTIFYING CORPORATE GOVERNANCE RISK FOR LATIN AMERICAN ADRS AND BANKS

Germán Creamer
Columbia University and Tulane University
Computer Science Department
500 West 120th. Street Rm. 450
New York, NY 10027
gcreamer@cs.columbia.edu

Yoav Freund
Columbia University
Center for Computational Learning Systems
500 West 120th. Street Rm. 529
New York, NY 10027
freund@cs.columbia.edu

ABSTRACT

The objective of this paper is to demonstrate how the boosting approach can be used to quantify the corporate governance risk in the case of Latin American markets. We compare our results using Adaboost with logistic regression, bagging, and random forests. We conduct ten-fold cross-validation experiments on one sample of Latin American American Depository Receipts (ADRs), and on another sample of Latin American banks. We find that if the dataset is uniform (similar types of companies and same source of information), as is the case with the Latin American ADRs dataset, the results of Adaboost are similar to the results of bagging and random forests. Only when the dataset shows significant non-uniformity does bagging improve the results. Additionally, the uniformity of the dataset affects the interpretability of the results. Using Adaboost, we were able to select an alternating decision tree (ADT) that explained the relationship between the corporate variables that determined performance and efficiency.

KEY WORDS
Corporate governance risk analysis, machine learning, Adaboost, data mining

1 INTRODUCTION

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

We expect the principal agent problem to have an important effect on company performance and efficiency. This is true in the case of the financial markets of countries in the process of development (emerging markets) because of their lax security regulations. The study of corporate governance in emerging markets is especially important because these markets have become increasingly 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 represent a certain number of shares of a foreign company in the major US stock markets such as the NYSE. We will concentrate only on Latin American ADRs and banks domiciled in Latin American countries.

In this article we demonstrate how the boosting approach can be used to evaluate the corporate governance risk. Here we create a predictive model 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 and of selected accounting ratios that are known to be important in evaluating corporate governance risk. We use Adaboost (Freund and Schapire 1997) as the learning algorithm of our predictive model. A second objective of this paper is to evaluate the use of Adaboost as a predictive and interpretative tool for corporate governance risk.

Previous studies on U.S. securities (see the pioneering works of Altman, 1968 and Beaver, 1966 and also see Altman (1974), (1989), (1998), Barr (1994), Collins and Green (1982), Chen and Lee (1995), Clarke and McDonald (1992), Goudie and Meekks (1992), Hudson (1997), Lane et al. (1986), Lau Hing-Ling (1987), Moyer (1977), Ohlson (1980), Pinches and Ming (1973), Queen and Roll (1987), Rose and Giroux (1984), and 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 which is usually assumed to be a linear system. A major limitation of this methodology is that non-linearities have to be incorporated manually.

In contrast, machine learning methods such as boosting and support vector machine avoid the question of modeling the underlying distribution and focus on making accurate predictions for some variables given others variables. Breiman (2001a) contrasts these two approaches as the data modeling culture and the algorithmic modeling culture. According to Breiman (2001a), 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. The main drawback of algorithmic modeling, according to Breiman, is that the generated models are hard to *interpret*.

This paper describes a first attempt in applying the algorithmic modeling approach to the analysis of corporate governance risk, comparing the results of Adaboost to logistic regression, random forest, and bagging, and evaluating its accuracy as well as its interpretability.

2 METHODS

2.1 LOGISTIC REGRESSION

The logistic regression models the posterior probabilities of L classes using linear regression. The model is a series of ordinary regressions where L-1 logit transformations or log-odds are the dependent variables:

$$\log\frac{Pr(C=1|X=x)}{Pr(C=L|X=x)} = \beta_{10} + \beta_1^T \text{ x}$$
$$\log\frac{Pr(C=2|X=x)}{Pr(C=L|X=x)} = \beta_{20} + \beta_2^T \text{ x}$$
...
$$\log\frac{Pr(C=L-1|X=x)}{Pr(C=L|X=x)} = \beta_{(L-1)0} + \beta_{L-1}^T \text{ x}$$

Taking the exponential of the log-odds, we can calculate the probabilities of each class as follows:

$$Pr(L=k|X=x)=\frac{e^{\beta_{l0}+\beta_l^T x}}{1+\sum_{r=1}^{L-1} e^{\beta_{r0}+\beta_r^T x}}, \text{ r = 1, ..., L-1}$$

The summation of these probabilities equal one. Logistic regression results are better interpreted using the odds ratios which can be computed by raising the e number to the power of the logistic coefficients (see Hatie, Tibishirani, and Friedman 2003).

2.2 BAGGING AND RANDOM FORESTS

Bagging was proposed by Breiman (1996) as a method that reduces the variance of a prediction function. Bagging or bootstrap aggregation averages the prediction of classifiers by the generation of different bootstrap samples. Each bootstrap sample is generated by obtaining uniform samples with replacement from the training set. Bagging has been shown to be particularly effective for reducing the variance of decision trees, according to Breiman et al. (1984).

Each sample S_i where $i = 1, .., n$ generates a classifier C_i using the initial prediction function. The final classifier $C*$ is obtained as an average of all the classifiers obtained from the bootstrap samples:

$$C* = 1/n \sum_{i=1}^{n}(C_i)$$

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

2.3 ADABOOST

Adaboost is a general discriminative learning algorithm invented by Freund and Schapire (1997).

$F_0(x) \equiv 0$
for $t = 1 \ldots T$
  $w_i^t = \exp(-y_i F_{t-1}(x_i))$
  Get h_t from *weak learner*
  $\alpha_t = \frac{1}{2}\ln\left(\frac{\sum_{i:h_t(x_i)=1,y_i=1} w_i^t}{\sum_{i:h_t(x_i)=1,y_i=-1} w_i^t}\right)$
  $F_{t+1} = F_t + \alpha_t h_t$

Figure 1. The Adaboost algorithm.

The basic idea of Adaboost is to repeatedly apply a simple learning algorithm, called the *weak* or *base* learner, to different weightings of the same training set. In its simplest form, Adaboost is intended for binary prediction problems where the training set consists of pairs $(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)$, x_i corresponds to the features of an example, and $y_i \in \{-1, +1\}$ is the binary label to be predicted. A *weighting* of the training examples is an assignment of a non-negative real value w_i to each example (x_i, y_i).

On iteration t of the boosting process, the weak learner is applied to the training set with a set of weights $w_1^t, \ldots, w_m^t$ and produces a prediction rule h_t that maps x to $\{0, 1\}$. The requirement on the weak learner is for $h_t(x)$ to have a small but significant correlation with the example labels y when measured using the *current weighting of the examples*. After the rule h_t is generated, the example weights are changed so that the weak predictions $h_t(x)$ and the labels y are decorrelated. The weak learner is then called with the new weights over the training examples, and the process repeats. Finally, all of the weak prediction rules are combined into a single *strong* rule using a weighted majority vote. One can prove that if the rules generated in the iterations are all slightly correlated with the label, then the strong rule will have a very high correlation with the label – in other words, it will predict the label very accurately.

The whole process can be seen as a variational method in which an approximation $F(x)$ is repeatedly changed by adding to it small corrections given by the weak prediction functions. In Figure 1, we describe Adaboost in these terms. We shall refer to $F(x)$ as the *prediction score* in the rest of the paper. The strong prediction rule learned by Adaboost is sign$(F(x))$.

2.4 ALTERNATING DECISION TREES

Similarly to Bagging, Adaboost is often used with a decision tree learning algorithms as the base learning algorithm. We use Adaboost both to learn the decision rules constituting the tree and to combine these rules through a weighted majority vote. The form of the generated decision rules is called an *alternating decision tree* (ADT) (Freund and Mason 1999).

We explain the structure of ADTs using one of the trees (see Figure 2) that we obtained using data from Latin American ADRs. The problem domain is corporate performance prediction, and the goal is to separate stocks with high and low value based on 17 different variables. The tree consists of alternating levels of ovals (*prediction nodes*) and rectangles (*splitter nodes*). The first number within the ovals defines contributions to the prediction score, and the second number indicates the number of instances. The dotted lines indicate where the tree has additional non relevant nodes. In this example, positive contributions are evidence of high performance, while negative contributions are evidence of corporate financial problems. To evaluate the prediction for a particular company we start at the top oval (0.042) and follow the arrows down. We follow *all* of the dotted arrows that emanate from prediction nodes, but we follow *only one* of the solid-line arrows emanating from a splitter node, corresponding to the answer (yes or no) to the condition stated in rectangle. We sum the values in all the prediction nodes that we reach. This sum represents the prediction score $F(x)$ above, and its sign is the final, or strong, prediction.

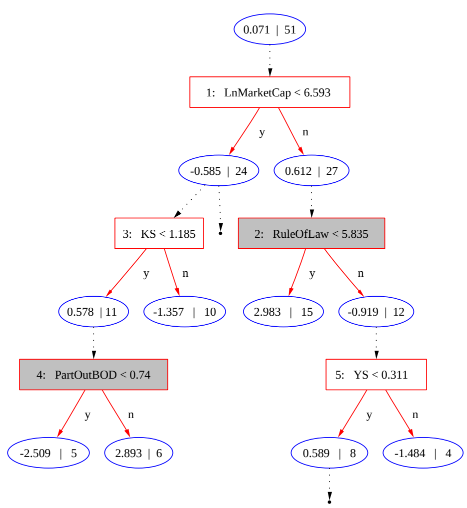

Figure 2. LAADR: representative adaptive tree. Corporate governance variables in gray.

For example, suppose we had a company for which LNMARKETCAP=6, KS=0.86, RULEOFLAW=7.02, and PARTOUTBOD=0.76. In this case, the prediction nodes that we reach in the tree are $0.042, -0.7181, 0.583$, and 1.027. Summing gives a score of 0.9339, i.e., a very confident indicator that the company has a high market value.

The ADT in the figure was generated by Adaboost from training data. In terms of Adaboost, each prediction node represents a weak prediction rule, and at every boosting iteration, a new splitter node together with its two prediction nodes is added to the model. The splitter node can be attached to any previous prediction node, not only leaf nodes. Each prediction node is associated with a weight α that contributes to the prediction score of every example reaching it. The weak hypothesis $h(x)$ is 1 for every example reaching the prediction node and 0 for all others. The number in front of the conditions in the splitter nodes of Figure 2 indicates the iteration number on which the node was added. In general, lower iteration numbers indicate that the decision rule is more important. We use this heuristic to analyze the ADTs and identify the most important factors in corporate performance.

3 PERFORMANCE AND EFFICIENCY

We use Tobin's Q as the measure of performance for ADRs.[1] Tobin's Q is the ratio of the market value of assets to the replacement cost of assets. This is a measure of the real value created by management. A higher value of Tobin's Q indicates that more value has been added or there is an expectation of greater future cash flow. Hence, the impact of management quality on performance is captured by Tobin's Q. Any difference of Tobin's Q from one indicates that the market perceives that the value of total assets is different from the value to replace their physical assets. So, the value of internal organization, management quality, or the expected agency costs may explain the difference. Values of Tobin's Q above one indicate that the market perceives the firm's internal organization as effective in leveraging the companies assets, while a Tobin's Q below one shows that the market expects high agency costs. We use a proxy for Tobin's Q as the ratio of book value of debt plus market value of common stocks and preferred stocks to total assets[2].

For the Latin American banks, we use an efficiency measure based on data envelopment analysis (DEA) instead of Tobin's Q because some of the banks under study are not public companies or participate in very illiquid markets.

DEA, using linear programming, builds a frontier selecting the "best practice" firms, obtains an efficient score, and recognizes overuse of inputs or underproduction of outputs. The study of efficiency must focus only on the relative levels of input in relation to output, which is technical efficiency. In other words, given a certain level of input, how to maximize output; or given defined levels of output, how to minimize input. Examples of this approach appear in the early nonparametric frontier models (Charnes, Cooper and Rhodes 1978) and in some of the early parametric frontier models such as Aigner, Lovell and Schnidt (1977).

The linear programming DEA envelopment problem is as follows:

[1] Tobin's Q is the preferred indicator of performance in corporate governance studies such as in La Porta et al. (2002) and Hermalin and Weisbach (1991).

[2] Peterson and Peterson (1996), Chung and Pruitt (1994), and Perfect and Wiles (1994) indicate that this 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. The application of this proxy may have even reduced more the sample of ADRs.

$\max \eta$

η, μ

subject to

$X\mu \leq x_o$

$\eta y_o \leq Y\mu$

$\mu \geq 0$

μ is an L by 1 intensity vector, X is an n by 1 input matrix, Y is an m by L output matrix, and x_i and y_i are the columns of the input and output matrix respectively.

η is a radial measure of technical efficiency. This version of DEA is output oriented. So, if a producer is able to expand its output vector according to the restrictions imposed by the "best practice", then $\eta > 1$. If a producer is already in the efficient frontier, η is 1. This version of DEA assumes constant returns to scale and was proposed by Charnes, Cooper, and Rhodes (1978) (see also Knox Lovell 1993).

4 CORPORATE GOVERNANCE

For the corporate governance variables, we include the presence of insider ownership, variables related to the structure of the board of directors (outsiders on the board of directors, the size of the board of directors, and the double role of CEO as chairman of the board of directors and manager), institutional ownership, and corporate governance indicators at the country level according to La Porta et al. (1998) (efficiency of the judicial system, rule of law, risk of expropriation, risk of contract repudiation, corruption, quality of accounting system, and legal system). We also consider the following accounting variables: the logarithm of market capitalization for ADRs, and an equity index per country as a proxy for size for Latin American banks[3]; long-term assets to sales ratio for ADRs and long-term assets to deposits for banks for their effect in the reduction of the agency conflict[4]; debt to total assets ratio as a capital structure indicator; operating expenses to sales ratio as an efficiency or agency cost indicator[5]; operating income to sales ratio as a market power proxy, and to indicate cash available from operations; and capital expenditures to long-term assets ratio[6] as a proxy for the relationship between growth and the possibility of investing in discretionary projects. We use region and sector as indicators of the geographical area and industrial sector in which the company operates (see Table 1).

We study insider ownership as a relevant variable because the separation of ownership and control is seen as an opportunity for managers to accumulate wealth at the expense of shareholders (Berle and Means 1932; Jensen and Meckling 1976; and Shleifer and Vishny 1997).

The next corporate governance variable that we explore is the presence of outsiders on the board of directors (for ADRs) or number of insiders on the board of directors (for banks). The board of directors plays a high-level counsel and control role in any organization. However, it is necessary that the board of directors include outsiders (who are not part of the management team) and maintain a minimal level of ownership to ensure their interest in the performance of the company. According to Jensen (1993), a board of directors may fail due to a strong emphasis on the CEO's personal agenda, a low equity ownership among the board of directors' members, an excessively large board of directors, and a culture of consent instead of dissent. Fama (1980) and Fama and Jensen (1983) explain how the separation between control and security ownership can be an efficient structure because professional outside directors may limit the power of agents to expropriate the residual claimants' interest. The size of the board of directors is also a relevant variable, according to Yermack (1996) and Fuerst and Kang (2000), because the size of the board of directors has an inverse association with firm value in the case of large US industrial corporations. Lipton and Lorsch (1992) and Jensen (1993) recommend that companies limit board membership to no more than seven or eight members. We also study the double role of the CEO as chairman of the board of directors. According to Jensen (1993), companies should separate the CEO role from the chairman role because of the need for independence. If the CEO is also chairman of the board, the dual role may have a negative impact on performance. Even more, Jensen recommends include active investors who hold large equity or debt position in a company and take part in their strategic decisions.

Institutional ownership is another mechanism to control managers' actions because of large institutional shareholders' roles as active monitors. Results might be ambiguous if there is insider ownership or hidden investment, because large shareholders may manage the firm for their own benefit only, and not for the benefit of the majority of small shareholders.

5 EXPERIMENTS

The data we used in our experiments are from Latin American ADRs (LAADR) and Latin American banks (LABANKS). These data are described in Appendix 1. We conducted a logistic regression with the corporate governance and accounting variables described in the previous section, using Tobin's Q and the DEA technical efficiency indicator as the dependent variables for LAADR and LABANKS respectively. The logistic regression includes a dummy variable for industrial sectors. We calculated the efficiency indicators for each country because of the differ-

[3] We used the equity index instead of equity value because efficiency is calculated country by country. We are interested in the effect of the relative size by country on efficiency instead of its absolute value.

[4] Assets can be monitored very easily and they can become collateral either for the development of new projects or to finance new acquisitions.

[5] 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.

[6] 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.

Table 1. Description of Variables

TobinQ	Tobin ' s Qwhich 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 stocks and preferred stocks to total assets
PartOutBOD	% outsiders on the board of directors
LnDIR	Natural logarithm of board size
InstPart	% institutional ownership
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.
ChairmanCEO	1 if CEO is chairman, 0 otherwise
LnMarketCap	Natural logarithm of market capitalization, used to measure firm size
KS or KD	The 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 either for the development of new projects or to finance new acquisitions.
YS	The ratio of operating income to sales
DebtRatio	The ratio of debt to total assets, used as a capital structure variable. Emerging markets 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.
Equity index	Index of equity according to country of residence. This is a measure of size applied to 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. Similarly, this efficiency ratio may also reveal agency costs or agency conflicts. 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. (This is different from the DEA technical efficiency indicator).
IK	The ratio of capital expenditures to long term assets (stocks of property, plant and equipment)
AvgParticipation	This is a measure of ownership concentration. This is calculated as the average of the participation of the three largest shareholders per firm
The following corporate governance variables are from La Porta et al. (1998):	
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)
French	If the firm is domiciled in a country that is part of the Napoleonic or French legal family according to La Porta et al.
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.
Corruption	Indicator of level of government corruption according to ICR. Low levels indicate higher corruption, such as solicitation of bribery by government officials
EfficiencyJudicialSystem	Index about the level of efficiency of the legal system according to the agency Business International Corp. Scale is from zero to ten. Lower values correspond to lower efficiency levels.
RiskOfExpropriation	Risk of confiscation or nationalization according to ICR. Scale is from zero to ten. Lower values imply higher risks.
RiskOfContractRepudiation	Risk of modification of a contract by economic, social or political reasons as defined by ICR. Lower values correspond to higher risks.
Accounting	La Porta et al. (1998: 1125) describes this item as: "Index created by examining and rating companies' 1990 annual reports on their inclusion or omission of 90 items. These items fall into seven categories (general information, income statements, balance sheets, fund flow statement, accounting standards, stock data and special items). A minimum of three companies in each country were studied."

ences between accounting systems in the countries under study. Hence, efficiency of banks is calculated in relation to their peers in their country.

For the logistic regression analysis and for all the learning algorithms, we eliminated variables that indicated multicollinearity. For LABANKS the variables eliminated were risk of contract repudiation, legal system, region, corruption, and debt ratio. For LAADR, the variables that we eliminated were risk of expropriation, risk of contract repudiation, and region. We also improved the results of the logistic regression by eliminating those variables whose inclusion increased the test error.

We used Adaboost to classify stocks above and below the median. In the LAADR sample, the median is very close to one. So, the results can be interpreted as the classification between those stocks with a market value of its assets above (Tobin's Q greater than one) or below (Tobin's Q smaller than one) its costs of replacement. For LABANKS, the classification is between more efficient and less efficient banks. The results of ADTs must be interpreted as companies with positive scores have high Tobin's Q or in the case of banks are efficient, while companies with negative scores have low Tobin's Q or are inefficients banks.

We performed tenfold cross-validation experiments with ten iterations to evaluate classification performance on held-out experiments using Adaboost. We used the MLJAVA package, which implements the alternating decision tree algorithm described in Freund and Mason (1999). [7]

[7]If interested in using MLJAVA, please contact `freund@cs.columbia.edu`

To evaluate the difficulty of the classification task, we compared our method, Adaboost, with random forests using the software Random Forests V5.0. [8] We run our experiments with 1000 trees and four variables randomly selected at each node in order to reduce the test error.

To check for the possibility that the Adaboost results could be improved because of the characteristic instability of Adaboost, we applied bagging to Adaboost (bagged boosting). We created ten folds for testing and training. We obtained 100 bootstrap replicates of each testing fold. We averaged the score of the bootstraps of each fold to get the estimated class. Finally, we averaged the test error of the ten folds. We also compared ADTs with a single tree classifier and with a stumps averaged classifier trained using boosting. We evaluated the differences between the average of the test error of Adaboost with the test errors of the rest of the learning algorithms using the t-test.

Using the results of Adaboost, we ranked the variables according to their frequency weighted by the position of the node. So variables that are added in earlier boosting iterations are considered more important.

For each sample, we selected the ADT that had the lowest test error, and had the most important ranked variables located in the same nodes that at least 60% of the trees. Finally, to confirm the structure of the ADTs, we run new regression tests for each sample with the variables that were selected as the most important variables for each ADT.

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

5.1 RESULTS

The evolution of the training and testing errors are in Figures 3 and 4. The single tree boosting behaves similarly to Adaboost, while the stumps boosting shows a poorer performance for LAADR, while it shows a better performance for LABANKS during the first 10 iterations.

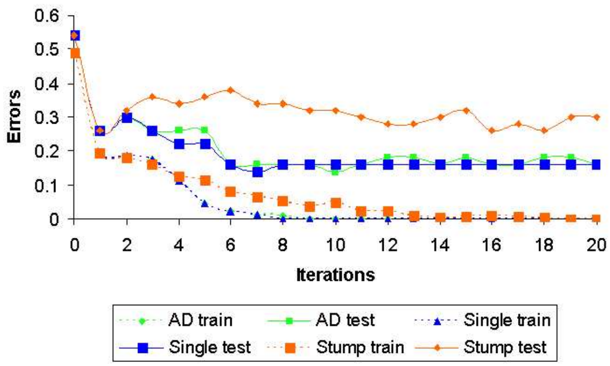

Figure 3. LAADR: training and testing error

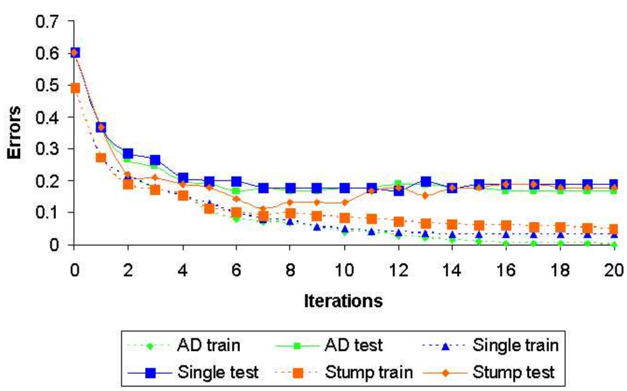

Figure 4. LABANKS: training and testing error

The ROC curve for LABANKS generated using Adaboost shows a larger proportion of true positives versus false positives in comparison to the LAADR case (see Figures 5).

The results of the testing errors for the learning algorithms used are shown in Table 2. As both our datasets are very small (51 examples in LAADR and 99 examples in LABANKS), evaluating the statistical significance of the different models and the comparison of their test errors is difficult. Acknowledging these limitations, we present the results of the t-test. The t-test indicated that there was a significant difference between the test errors of Adaboost and random forests for LAADR. There were no differences of the test errors for the rest of the tests in both samples.

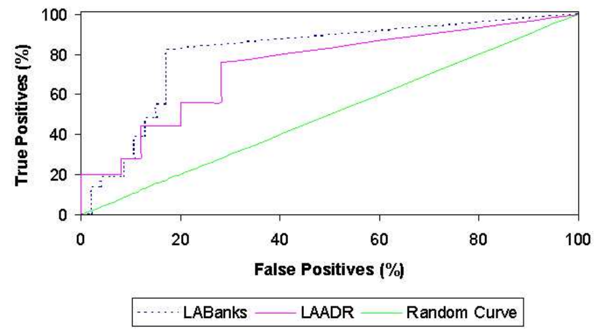

Figure 5. LAADR and LABANKS: ROC curve

Table 2. Testing errors of learning algorithms

	LAADR		LABANKS	
	Test error	St. dev.	Test error	St. dev.
Adaboost	14.0%	16.5%	17.8%	9.4%
Single tree	16.0%	12.7%	17.8%	11.9%
Stumps	32.0%	19.3%	13.3%	11.5%
Bagged boosting	22.0%	23.9%	13.33%	8.66%
Random forests	32.0% *	16.9%	16.67%	18%
Logistic regression	23.7%	18.5%	20.1%	13.3%
No observations	51		104	

*: 5% significance level of t-test difference between test error and Adaboost test error

Table 4 indicate the importance of each variable according to Adaboost and random forests. The results of both algorithms coincide in terms of what the most important variables are. Four top variables of the LAADR dataset according to Adaboost are ranked between the top six variables in random forests. In the case of LABANKS, four variables chosen by Adaboost are ranked between the top five variables according to random forests. Considering the similarity of the most important variables selected by random forests and Adaboost, we discuss the ADTs for LAADR and LABANKS; however, we do not discuss in detail the results of random forests to avoid repetition. The results of bagged boosting cannnot be interpreted in terms of the impact of each variable on performance and efficiency because of the large number of trees generated.

The odds ratios of logistic regression also confirm the importance established by Adaboost and random forests of the following variables: long-term assets to sales ratio and rule of law for LAADRs; and long-term assets to deposits ratio, insider ownership, and risk of expropiation for LABANKS.

We also run a linear regression between Tobin's Q and the main variables of LAADR, and between the technical efficiency indicator and the main variables of LABANKS (see Table 3) in order to evaluate the structure of the ADTs. Companies characterized by a logarithm of market capitalization greater than 6.73 present an inverse relationship between the rule of law indicator and Tobin's Q, and between operating income to sales ratio and Tobin's Q. Smaller companies have an inverse relationship between long-term assets to sales ratio and Tobin's Q, while they have a direct relationship between participation of outsiders in the board of directors and Tobin's Q as the ADT indicates. Likewise, Latin American banks with an equity index smaller than 0.75 show an inverse relationship between long-term assets to deposits ratio and technical efficiency. Banks with an equity index smaller than 0.13 show that low country risk of expropiation is associated with a higher efficiency level, and a long-term assets to deposits ratio smaller than 0.076 has an inverse relationship with efficiency.[9] The regression analysis confirms the structure of the ADTs for LAADR and LABANKS.

Table 3. Regression of relevant variables for ADTs. Tobin's Q is dependent variable for LAADR and technical efficiency indicator is dependent variable for LABANKS

	LAADR				LABANKS					
	Ln Mkt Cap.>=6.7		Ln Mkt Cap.<6.		Equity < 0.75		Equity < 0.13		KD < 0.076	
	Coeff.	t stat.	Coeff.	t stat.	Coeff.	t stat.	Coeff.	t stat.	Coeff.	t stat.
Rule of Law	-0.59 **	-2.41								
YS (Oper. income / sales)	-4.55 **	-2.52								
Equity										
KS (L.T. assets/sales)			-0.14 **	-2.17	20.72 ***	6.39			4.64	0.78
Partic. Outsiders BOD			0.37	1.11						
Risk of Expropiation							-0.2262	-0.21		
R square	0.29		0.245		0.3351		0.001		0.0079	

Note: Significance level: **: 5%, ***: 1%. Corporate governance variables in gray.

6 METHODOLOGICAL FINDINGS

The tenfold LAADR test errors do not show any significant difference between Adaboost and the other learning algorithms according to the t-test, with the exception of random forests, which shows a higher test error of 32%. For the tenfold LABANKS cross validation, Adaboost has a 17.8% test error. Bagged boosting reduces the Adaboost test error to 13.33%.

It seems that the advantage of using bagging over Adaboost depends on the uniformity of the dataset. LAADR is a more uniform sample than LABANKS. LAADR only includes companies of large Latin American countries that fully obey the registration requirements of the SEC, including complying with US GAAP, while LABANKS includes banks of different size and following different accounting standards of Latin American countries. If the dataset is an agglomeration of several different datasets, such as in LABANKS, bagging can improve the results; however if the dataset is uniform such as in LAADR, bagging or random forests do not show any improvement over Adaboost. Therefore, stability is not a property that only depends on the learning algorithm; it also depends of the uniformity of the dataset.

The logistic regression analysis offered some insight about the relevance of the most important variables, however it was not possible to capture the interaction of these variables with the limited amount of data that we had. In contrast, Adaboost helped to rank the variables according to their importance, and also modeled their interaction.

In synthesis, Adaboost performed in a similar way to other learning algorithms such as bagging and random forests, and had the capacity to interpret the results because of the limited number of trees that were generated in contrast to the requirements of the other methods.

7 FINANCIAL INTERPRETATION

Comparing the ADTs of LAADR and LABANKS (see Figures 2 and 6), the main distinctive variable is the size of the company measured by the logarithm of market capitalization for ADRs and equity index for LABANKS. This result coincides with the previous study of Fama and French (1992) in USA, which indicated that size is a key factor to explain the rate of return of stocks. ADRs and banks around or above the median perform better than the rest. Large companies in emerging markets are likely to be oligopolies or monopolies in their area of activity. The efficiency of smaller banks is also affected when there is a high country risk of expropiation. However, the performance of LAADR improves in countries with a weak rule of law. Large Latin American companies probably perform better in environments with a weak tradition of law and order because of the close family relationships that help them to influence government decisions in their favor. The benefits of these government private sector connections seem to be less important for the small banks sector.

In countries with a strong rule of law and order, large companies may still have an important agency conflict that affects their performance if the cash available for operations is too high, as a large operating income to sales ratio indicates. An excessive amount of cash may allow managers to spend it on projects that benefit them directly instead of increasing the value of their companies. A large operating expenses to sales ratio may also indicate an agency conflict. Among the medium and large companies, 58% have an excessive efficiency ratio in relation to the threshold level found by Adaboosting.

The performance of small and medium size companies improves if the proportion of long-term assets to sales is below 0.97 for LAADR companies. For Latin American banks, the efficiency improves when the long-term assets to deposits ratio is below 0.076 (close to the median). These indicators are important to reveal 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.

[9] Table 3 shows results of a linear regression for LABANKS where the dependent variable is the technical effi ciency indicator. A higher value of this indicator implies less effi ciency. In contrast, a higher value of rule of law and risk of expropiation indicators imply a better indicator.

Table 4. LAADR and LABANKS: Statistics, logistic regression, Adaboost, and random forest results relating corporate governance variables, Tobin's Q, and efficiency.

	LAADR									LABanks									US
	Statistics				Logit	Adaboost		RF		Statistics				Logit	Adaboosting		RF		Stat.
	Q25	Median	Q75	Mean	Odds ratios	Threshold avg.	Rank	z-score	Rank	Q25	Median	Q75	Mean	Odds ratios	Threshold avg.	Ranks	z-score	Rank	Mean
LnMarketCap (Nat. log market capitalization)	5.44	6.73	7.49	6.57	0.000	6.7341	1	26	1	(Not used)									
Equity index	(Not used)									0.04	0.16	0.50	0.30	0.0	0.72879	2	34.71	1	
Equity index (2nd. Node)	(Not used)														0.11072				
IK (Capital expenditures/ long-term assets)	0.05	0.08	0.13	0.10			10	6	3	(Not used)									
EfficiencyJudicialSystem (Effic. legal system)	6.00	6.00	7.25	6.50			9			6.00	6.25	6.75	6.43	0.456		9	11.54	4	10
RuleOfLaw (Law and order tradition)	5.35	5.35	7.02	5.82	0.027	5.8378	2	6	2	2.50	6.32	6.67	5.27	0.6335		7	7.70	6	10
PartOutBOD (% outsiders as directors)	60.0%	77.0%	87.0%	68.1%	0.000	0.7438	5	2	6	75%	94%	100%	85%	2.5271		7	3.03	9	60.0%
Avg Participation	(Not used)									50.0%	92.7%	100.0%	74.5%	0.1566		6	4.14	8	
Efficiency (Operating expenses / sales)	0.10	0.16	0.23	0.16		0.2088	6	5	4	(Not used)									
YS (Operating income / sales)	0.13	0.23	0.35	0.25	0.000	0.3122	3			(Not used)									
Debtratio(Debt / total assets)	0.46	0.59	0.80	0.61				2	5	0.89	0.92	3.23	77.96						
LnDir (Natural log number directors)	1.95	2.20	2.30	2.08	1.04E+11		7			1.792	2.1972	2.3979	2.10	4.05E-01		5	12.37	3	
InstPart (% institutional equity ownership)	15.0%	44.0%	71.0%	43.2%	27.7571		8												
Corruption (Level of government corruption)	4.77	5.30	5.30	5.21	82222.46					5.00	5.18	6.02	5.45						8.63
RiskOfExpropiation (Risk confiscation)	6.95	7.29	7.50	7.09						5.91	6.57	7.50	6.67	0.4704	7.035	3	10	5	9.98
KS (L.T. assets/sales for ADRs) or KD (LT ass./deposits for banks)	0.73	1.44	2.20	1.82	53157	1.0032	4			0.04	0.06	0.10	0.11	2.E+12	0.082	1	33	2	
KS or KD (2nd. Node)	(Not used)														0.11406				
T_Insiders (% insider's equity ownership)	0.0%	0.0%	2.0%	10.4%				0	7	0.0%	0.0%	1.2%	8.8%	0.5299		4	4.18	7	
No. Directors	7	9	10	8.29						6.00	9.00	11.00	9.59						9.35
RiskOfContractRepudiation (Contract change)	6.30	6.55	6.80	6.33						4.91	5.18	6.30	5.68						9
TobinQ (Tobin's Q: performance)	0.91	1.04	1.44	1.38						(Not used)									

Note: US number of directors, and US average of outside directors according to Denis and Sarin (1999). US CEOs who are also chairman, and insiders' ownership from Fuerst and Kang (2000).
Country corporate governance variables from La Porta et al. (1998). RF: Random forests. Q25: 25th. percentile. Q75: 75th percentile. Logistic regression includes dummy variables to control for sector.
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.

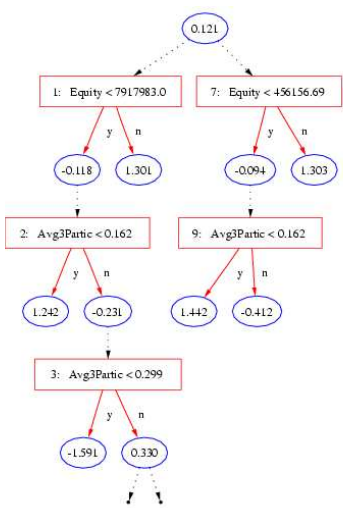

Figure 6. LABANKS: representative adaptive tree

According to the ADT for LAADR, the composition of the board of directors is important for smaller companies which have a capital sales ratio below 0.97. In these cases, the participation of outsiders on the board of directors above a level of 75.5% is a relevant factor to improved performance. The finance literature indicates that outside directors supervise managers (Weisbach 1988 and Shivdasani 1993). Weisbach (1988) finds that outsider-dominated boards are more likely to remove CEOs than firms with insider-dominated boards, especially when firms show poor performance.[10] Denis and Sarin (1999) find that companies that increase the proportion of outsiders on the board of directors or reduce ownership concentration have above average returns in the previous year. However, Yermack (1996), MacAvoy et al. (1983), Hermalin and Weisbach (1991), and Bhagat and Black (1999 and 2000) find little correlation between composition of board of directors and performance. One possible explanation for these results is that the CEO hires outside directors; hence, directors do not dissent (see Crystal 1991). This hypothesis is reinforced by Core et al. (1999), who find that CEO compensation is a decreasing function of the share of inside directors, and is an increasing function of the share of outside directors chosen by the CEO.

Inside directors also play an important role in the board of directors for strategic planning decisions, reviewing functional performance by areas and, in some cases, evaluating if there are important differences between the CEO's perspective and what is happening in the firm on a daily basis. [11] Baysinger and Butler (1985) propose that an

[10] In the case of Italy this situation is different. Volpin (2002) finds that the probability of turnover and its relationship to performance is lower for executives who are part of the family of the controlling shareholder. Rosenstein and Wyatt (1990) find that announcements of outside directors are related to positive excess returns.

[11] Klein (1998) found that inside director participation in investment

optimal board of directors should have a combination of inside, independent, and also affiliated directors. Baghat and Black (2000) suggest that boards should not be composed only of independent directors because of their findings that board independence does not improve performance, and because inside directors may bring the additional benefits explained above. This may explain why the ADT suggests that the maximum participation of outsiders in the board of directors of LAADR companies should be 75.5%.

Insider ownership does not seem to affect the performance of LAADR companies. This can be understood in terms of the information published in the proxy statements of ADRs. These reports are not under the same strict control that the financial statements are. As a result, it is possible that many firms did not include relevant information about managers, ownership structure and board composition due to the need to protect shareholders against potential kidnapping, assault, etc. Hence, only the major shareholders are registered. In the case of LABANKS, according to logistic regression and Adaboost, insider ownership is the third and fourth most important variable in explaining efficiency.

Management with a high level of ownership are likely to be able to steer corporate decisions toward their own interests at the expense of corporate interests. This could be the case of strong family groups that control a company. These family groups may use their great bargaining power to make corporate decisions that benefit companies where they have a great interest. For example, banks may direct an important part of their loan portfolio to companies where managers or insiders have a significant interest. 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 nineties. Jensen and Meckling (1976) in their classic work described this behavior where large investors as equity holders will benefit when the firm takes an excessive risk because of the potential benefit on the upside, while the other stakeholders, such as the creditors, bear all the risk. Hermalin and Weisbach (1991) had already proposed that agency costs increase with ownership, such as in the case of family firms. La Porta el al. (1999) also mention that the agency problem in these companies is that the dominant family owner-manager may expropriate minority shareholders. Hence, there is a strong incentive to be a large shareholder in developing countries.

The limited impact of size of board of directors, the double role of CEO as manager and chairman of the board of directors, and composition of board of directors (percent of outsiders) on performance and efficiency (in the case of LABANKS) using logistic regression or ADAboost are findings similar to what previous studies have indicated in USA.[12]

8 FINAL COMMENTS AND FUTURE WORK

Comparative regional studies always have a major problem in terms of how to integrate data coming from different sources, and generally with different standards. We saw that this problem was implicit in the LABANKS dataset. We think that this research can be improved by enlarging the dataset and running the learning algorithms in subsets aggregated by regions or corporate governance systems.

This paper shows that Adaboost performs similarly to logistic regression, random forests, and bagging with stable datasets, and the structure of the ADTs are confirmed by the regression analysis. Additionally, we show how Adaboost can be used as an interpretative tool to evaluate the impact of corporate governance risk factors on performance and efficiency.

Initially, the corporate governance variables do not seem to be very relevant to predicting corporate performance. However, when the results of these variables were interpreted together with the accounting variables using ADTs, the effect of corporate governance on performance became evident. The recent cases of US bankruptcies have demonstrated that when companies are doing very well, the corporate governance variables do not seem to be relevant. However, in moments of financial distress, corporate governance variables play a very important role in improving performance and efficiency.

Acknowledgements

GC thanks Paul Spindt, Tom Noe, Kenneth Jameson, Greg Buchholz, John M. Trapani, participants in the 2003 Latin American Studies Association meeting, 2000 Eastern Finance Association meeting, and 2001 Tulane University Latin American Research Consortium annual meeting for their helpful comments on partial versions of this paper. GC also thanks the institutional support of the Board of Regents of the State of Lousiana, and 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 authors.

References

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

Altman, E. I. 1968. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*. 23 (4).

Altman, E. I. 1974. Evaluation of a Company as a Going Concern. *The Journal of Accountancy*.

Altman, E. I. 1989. Measuring Corporate Bond Mortality and Performance. *The Journal of Finance*. 54.

Altman, E. I. 1998. Credit-risk Measurement and Management: the ironic challenge in the next decade. *Financial Analysis Journal*.

Barr, R., L. M. Seiford and F.Thomas. 1994. Forecasting Bank Failure: a non-parametric frontier estimation approach. *Recherches Economiques de Louvain* 60(4).

Baysinger, Barry and Henry Butler. 1985. Corporate governance and the board of directors: performance effects of changes in board composition. *Journal of Law, Economics and Organization* 1, 101-124.

Beaver, W., 1966, Financial Ratios as Predictors of Failure *Journal of Accounting Research*, 4, 71-111.

Berle, A., and G. Means. 1932. *The Modern Corporation and Private Property*. New York: Harcourt, Brace and World.

Bhagat, S., and B. Black. 2000. Board independence and long-term firm performance. Working Paper. Stanford University.

———. 1999. The uncertain relationship between board composition and firm performance. *Business Lawyer* 54, 921-963.

Breiman, L. 1996. Bagging predictors. *Machine Learning* 24, 123-140.

———. 2001a. Statistical modeling: The two cultures. *Statistical Science* 16, 199-231.

———. 2001b. Random forests. Technical report, University of California, Berkeley.

Breiman, L., J. H. Friedman, R. Olshen and C. Stone. 1984. *Classification and Regression Trees*. Belmont: Wadsworth & Brooks.

Byrne, J. A. 1996. The best and worst boards. *Business Week* November 25.

Charnes, A., W.W. Cooper,, and E. Rhodes. 1978. Measuring the Efficiency of Decision Making Units, *European Journal of Operational Research* 2. 429-44.

Chen, K. C.W. and Jevons Lee C.W. 1995. Accounting Measures of Business Performance and Tobin's q Theory. *Journal of Accounting, Auditing and Finance*.

Chung, K.H. and S. W. Pruitt. 1994. A Simple Approximation to Tobin's Q. *Financial Management* 23 (3), 70-74.

committees correlates with better firm performance.

[12]Bhagat and Black 2000 do not find that board independence leads 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.

Clarke, D. G. and J. B. McDonald. 1992. Generalized Bankruptcy Models Applied to Predicting Consumer Credit Behavior. *Journal of Economics and Business*.

Collins, R. A. and R. D. Green. 1982. Statistical Methods for Bankruptcy Forecasting. *Journal of Economics and Business*. 34, 349-354.

Core, J., R. Holthausen, and D. Larcker. 1999. Corporate governance, chief executive officer compensation, and firm performance. *Journal of Financial Economics* 51, 371-406.

Crystal, G. 1991. *In Search of Excess: The overcompensation of American executives*. New York: W. W. Norton Company.

Denis, D. J. and A. Sarin. 1999. Ownership and Board Structures in Publicly Traded Corporations. *Journal of Financial Economics* 52, 187-223.

Fama, E. 1980. Agency Problems and the Theory of the Firm. *Journal of Political Economy* 88, 288-307.

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

Fama, E., and K. French. 1992. The cross-section of expected stock returns. *Journal of Finance* 47 (2), 427-465.

Freund, Y., and R. E. Schapire. 1997. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences* 55 (1), 119-139.

Freund, Y., and Ll. Mason. 1999. The alternating decision tree learning algorithm. *Proceedings of the Sixteenth International Conference on Machine Learning*, 124–133.

Fuerst, O., and S. Kang. 2000. Corporate Governance, Expected Operating Performance, and Pricing. Working Paper. Yale School of Management.

Goudie, A.W. and G. Meeks. 1992. "Forecasting and Explaining Corporate Failure: A Study of Contingent Financial Statements". University of Cambridge, Department of Applied Economics.

Hastie, T., R. Tibishirani, and J. Friedman. 2003. *The Elements of Statistical Learning*. New York: Springer.

Hermalin, B., and M.S. Weisbach. 1991. The Effects of Board composition and Direct Incentives on Firm Performance. *Financial Management* 20, 101-112.

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., and W. H. Meckling. 1976. Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure. *Journal of Financial Economics* 3, 305-360.

Klein, April. 1998. Firm Performance and Board Committee Structure. *Journal of Law and Economics* 41, 275-303.

Knox Lovell, C.A. 1993. Production Frontiers and Productive Efficiency. in H. Fried, C.A. Knox Lovell, and S. Schmidt. 1993. *The Measurement of Productive Efficiency*. New York: Oxford University Press.

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

—— 1998. Law and finance. *Journal of Political Economy* 196, 1113-1155.

Lane, W. R., S. W. Looney and J. W. Wansley. 1986. An Application of The Cox Proportional Hazards Model to Bank Failure. *Journal of Banking and Finance*. 10, 511-531.

Lau Hing-Ling, Amy. 1987. A Five-State Financial Distress Prediction Model. *Journal of Accounting Research*. 25 (1).

Lindenberg, E. B., and S. A. Ross. 1981. Tobin's Q Ratio and Industrial Organization. *Journal of Business* 54 (1), 1-32.

Lipton, Martin, and Jay Lorsch. 1992. A Modest Proposal for Improved Corporate Governance, *Business Lawyer* 48, 59-77.

MacAvoy, P., S. Cantor, J. Dana, and S. Peck. 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*, Exhibit C.

Moyer, C. R. 1977. Forecasting Financial Failure: a re-examination. *Financial Management*.

Ohlson, J. 1980. Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*. 18 (1).

Perfect, S. B. and K. W. Wiles. 1994. Alternative Construction of Tobin's Q: An empirical comparison. *Journal of Empirical Finance* 1 (3/4), 313-41.

Peterson. P., and D. Peterson. 1996. Company Performance and Measures of Value Added. Charlottesville: *The Research Foundation of the Institute of Chartered Financial Analysts*.

Rosenstein, S., and J.G. Wyatt. 1990. Outside Directors, Board Independence, and Shareholder Wealth. *Journal of Financial Economics* 26,175-191.

Pinches, G. E. and K.A. Mingo. 1973. A Multivariate Analysis of Industrial Bond Ratings. *The Journal of Finance*. 28 (1).

Queen, M. and R. Roll. 1987. Firm Mortality: using market indicators to predict survival. *Financial Analysts Journal*.

Rose, P. S. and G. A. Giroux. 1984. Predicting Corporate Bankruptcy: an analytical and empirical evaluation. *Review of Business and Economics Research*. 19 (2).

Shleifer, A., and R. Vishny. 1997. A survey of corporate governance. *Journal of Finance* 2, 737-783.

Shivdasani, A. 1993. Board Composition, Ownership Structure, and Hostile Takeovers. *Journal of Accounting and Economics*, 16: 167-98.

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.

Yermack, D. 1996. Higher Market Valuation of Companies with a Small Board of Directors. *Journal of Financial Economics* 40, 185-211.

Zavgren, C. V. 1983. The Prediction of Corporate Failure: the state of the art. *Journal of Accounting Literature*. 2.

Appendix 1: Data

We used a sample of 51 stocks domiciled in Latin America (LAADR) (Argentina, Brazil, Chile, Colombia, Peru, Mexico, and Venezuela) that have issued ADRs of level II and III for the year 1998. Level I ADR are least restricted in their required compliance with US regulations, so we have not included them in our 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 complying 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 a US public company, and are therefore the most regulated. We chose ADRs from countries on the list of emerging markets database (EMDB) of the International Finance Corporation (IFC). [13].

We obtained the financial information from COMPUSTAT for the year 1998. The information on the value of market capitalization is from CRSP, and is compared with information from the NYSE. We obtained corporate governance information – such as list 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 ownership of a company by the CEO, managers, or relatives of the CEO, and members of the board of directors.

We also used a list of 99 Latin American banks, called LABANKS. LABANKS consists of banks domiciled in Argentina, Brazil, Chile, Colombia, Peru, Ecuador, and Bolivia[14] representing about 80% of the total assets of the private sector in the major Latin American countries.

We obtained the banks' corporate information from Internet Securities Inc., central bank, regulator and company websites. We collected financial as well as corporate information similar to that collected for ADRs. Our sample of banks is restricted by the availability of corporate finance information.

Most of the financial information is from 2000. Few companies that were merged or disappeared in 1998, were included using the financial statements of 1997. The corporate information is gathered from the period 1998-2000. Considering that the information about ownership structure is relatively stable, we do not foresee any major consistency problem.

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

[14]We were not able to include Venezuela's banks because the President of the Venezuelan Banking Association declined to supply any information to our research team and asked member banks not to supply any corporate information to us due to the increased risk of kidnapping that its members would be subject to if this information were distributed.