Lancaster University
MANAGEMENT SCHOOL

Lancaster University Management School
Working Paper
2007/008

Statistical modelling to predict corporate default for Brazilian companies in the context of Basel II using a new set of financial ratios

Joao Minussi, Didier Soopramanien and Dave Worthington

The Department of Management Science
Lancaster University Management School
Lancaster LA1 4YX
UK

The LUMS Working Papers series can be accessed at http://www.lums.lancs.ac.uk/publications/
LUMS home page: http://www.lums.lancs.ac.uk/

STATISTICAL MODELLING TO PREDICT CORPORATE DEFAULT FOR BRAZILIAN COMPANIES IN THE CONTEXT OF BASEL II USING A NEW SET OF FINANCIAL RATIOS

ABSTRACT

This paper deals with statistical modelling to predict failure of Brazilian companies in the light of the Basel II definition of default using a new set of explanatory variables.

A rearrangement in the official format of the Balance Sheet is put forward. From this rearrangement a framework of complementary non-conventional ratios is proposed.

Initially, a model using 22 traditional ratios is constructed. Problems associated with multicollinearity were found in this model. Adding a group of 6 non-conventional ratios alongside traditional ratios improves the model substantially.

The main findings in this study are: (a) logistic regression performs well in the context of Basel II, yielding a sound model applicable in the decision making process; (b) the complementary list of financial ratios plays a critical role in the model proposed; (c) the variables selected in the model show that when current assets and current liabilities are split into two sub-groups - financial and operational - they are more effective in explaining default than the traditional ratios associated with liquidity; and (d) those variables also indicate that high interest rates in Brazil adversely affect the performance of those companies which have a higher dependency on borrowing.

KEYWORDS: Default prediction, statistical modelling, non-conventional financial ratios, Basel II, Brazilian context.

1. Introduction

The environment of credit and the circumstances of lending have changed substantially in the last two decades. Traditional bankers and their managers, who intuitively believe in their feelings and experience to judge good and bad customers, are becoming consigned to history. Likewise, the traditional accounting analysts are embracing change and are being replaced by analysts with solid knowledge not only of accounting and finance, but also of other related areas, such as statistics, forecasting, data mining and econometrics. Several factors are responsible for these changes, but the rapid evolution in computer applications and the unprecedented levels of bankruptcy seen in the first years of this century are the most important points. The Bank for International Settlements – BIS (2003) has set new rules to increase the operational security of the banks. Banks are a major funding force behind entrepreneurship and sustainable economic development. When they are robust, capitalised and well prepared to assess risks correctly they will be more resilient to face periods of economic turmoil. According to McQuown (2003), the New Basel Capital Accord (Basel II) will allow banks to reduce unexpected losses, improve profitability, increase risk-carrying capacity and undertake more originations. Basel II lays down guidelines that all banks should develop systematic/validated methods for assessing the risks associated with business loans. In other words, the accord states that banks must have a robust system in place to validate the accuracy of estimated probabilities of default.

It has to be highlighted here that Basel II does not specify what type of model should be adopted by banks. However, when 'validated models' are mentioned, it is realistic to presume that credit scoring models will increasingly replace the conventional and more costly credit analysis which is based on subjective judgment. Credit scoring models employ quantitative methods for evaluating and predicting the credit worthiness of borrowers and should provide the foundation for the new financial environment, permitting banks to manage credit risk more effectively over the coming years. The adoption of statistical models is expected to increase, because banks that have a sound credit risk model in place will be allowed to set aside a lower amount of capital against loans issued to their low risk customers.

Another implication of Basel II is that validated models should be applied worldwide, whereas the preponderance of studies relating to bankruptcy forecasting has been carried out in developed countries, especially in the USA and Europe. Basel II also brings its own definition of defaulting companies, which differs from the definition used in most studies, i.e. companies which go into bankruptcy.

Furthermore recent research into predicting company bankruptcy has typically only employed publicly available datasets and has concentrated on developing and evaluating new modelling methods. There is less research that involves collaboration with banks where there is the opportunity to challenge the choice of explanatory variables used to predict risk of default. However, Basel II implies that banks should devise their own methods, as long as they are validated, which opens the door for research into new variables as well as into new models.

The main contributions of this paper are:

(i) To show the results of building a model for assessing credit risk consistent with the Basel II requirements from the perspective of a large Brazilian retail bank;

(ii) To evaluate the contribution of a new framework of financial ratios alongside those commonly used in predicting the risk of default of companies.

The remainder of this paper is organised as follows: Section 2 presents a brief literature review. In Section 3, the data set and methodology are described, including a proposed new framework of financial ratios. Section 4 describes how logistic regression models are used to predict the risk of default, and section 5 presents the main results of the empirical analysis using conventional financial ratios on their own and then alongside the proposed non-conventional ratios. In section 6 we discuss the implications of our results, draws conclusions and highlight areas for further research.

2. Brief Review of Literature

Over the last four decades, there has been increasing research into predicting credit risk and modelling of bankruptcy. Interest in this topic has even extended into other academic fields and many publications can be found in journals from other academic areas, such as in statistics, computation, artificial intelligence and operational research. In the related literature, different approaches and models for assessing probability of default are suggested. This brief review restricts itself to the application of statistical models for predicting the risk of default. More detailed reviews can be found in Balcaen and Ooghe (2004a), Balcaen and Ooghe (2004b) and in Aziz and Dar (2004).

Beaver (1967a), applying univariate analysis, is recognised as attributed as the pioneer in employing statistical methods using financial ratios to evaluate a firm's risk of defaulting. Shortly afterwards, Altman (1968) wrote a seminal paper in this area introducing multivariate analysis. His model is commonly referred to as "Z-score" and uses multiple discriminant analysis (MDA). He constructed a model which resulted in a discriminant function composed of 5 financial ratios using a sample of 66 companies. MDA was the dominant method for more than one decade. Ohlson (1980) is recognised as the pioneer in applying logistic regression based on financial ratios in bankruptcy studies. He criticised MDA for methodological reasons and others such as Zmijewski (1984), Zavgren (1985), Lo (1986), Funning and Coger (1994) and Lennox (1999) tend to agree with Ohlson (1980). Their main criticisms of discriminant analysis are the assumptions of normality of the independent variables and that the group dispersion (variance-covariance) is equal across time. Later researchers have used probit and logit methods more intensively, which require less restrictive assumptions on the independent variables.

Balcaen and Ooghe (2004a) provide a valuable overview and discussion of the application of “classical cross-sectional statistical methods”. Whilst noting that conditional probability models, e.g. logistic regression, make less restrictive assumptions than MDA, they emphasise that all classical cross-sectional statistical methods suffer from potential weaknesses associated with: (i) need for a dichotomous dependent variable, (ii) bias associated with the sampling method, (iii) non-stationary and unstable data, (iv) limited nature and accuracy of annual account information, (v) selection of independent variables, and (vi) the time dimension.

From the 90s, non-traditional statistical methods such as Artificial Neural Networks (ANNs) have been employed for credit risk modelling. Whilst achieving some impressive results, research is inconclusive about their benefits. For example Laitinen and Kankaanpaa (1999) compared six failure prediction techniques in terms of failure prediction accuracy for one data set, concluding that no superior method was found. In a comprehensive account of a range of alternative techniques and a literature-based comparison of empirical studies, Balcaen and Ooghe (2004b) found that on balance the more sophisticated alternatives techniques were of questionable benefit. Similarly O’Leary (1998) analysed 15 studies which made use of ANNs to predict corporate bankruptcy. However because they differed, for example in terms of datasets, software, variables, training and testing samples, they concluded that it was difficult to fairly compare their results.

One of the well-known practical weaknesses of the non-traditional statistical models is the problem of interpreting the resulting model. Balcaen and Ooghe (2004a) note that one of the advantages of logistic regression is that the coefficients can be interpreted separately reflecting the importance of each of the independent variables in predicting the estimated failure probability. However this is only the case when the independent variables do not suffer from multi-collinearity. Where multi-collinearity exists, logistic regression models can produce coefficients with the 'wrong', counter-intuitive sign.

Aziz and Dar (2004) analysed 89 studies of prediction models dealing with corporate bankruptcies including classical statistical methods, neural networks and decision trees. One of the issues they identify for all types of studies is that of small samples and the associated problem of model over-fitting. All but 6 of the 89 studies used datasets where the number of observations was less than 600. Furthermore, less than half the studies which they include used a hold-out sample, which is the recommended approach to reduce the risks of over-fitting. These authors also noted that previous studies have largely employed traditional financial ratios as explanatory variables, and that information on cash flow, basic firm characteristics, quality of management and market variables etc has played a relatively minor role in the task of prediction. Researchers who have investigated one or more of the above types of variables alongside traditional financial ratios include Ohlson (1980), Zavgren (1983), Keasey & Watson (1987), Sheppard (1994), Lussier (1995), Schumway (2001), Yu Du (2003), Lehmann (2003) Becchetti & Sierra (2003), Charitou *et al.* (2004) and Back (2005). Our research differs from these past studies insofar as we use traditional financial ratios and then propose a complementary framework of financial ratios not adopted in previous research relating to credit risk models.

As discussed in section 1, we are interested in the feasibility and benefit of applying statistical models of credit risk consistent with Basel II, from the perspective of a major Brazilian bank. Secondly we are interested in the benefits of introducing non-traditional financial ratios, as may well be possible when a ‘bank-specific’ model is required. Whilst these two issues could be researched via a number of the modelling approaches discussed above, logistic regression has been selected for the remainder of this paper. The advantages of this approach are first that logistic regression clearly shows the nature of the results, second that the relative transparency of logistic regression models is desirable for practitioners, and third that the same transparency also facilitates interpretation of the research results.

3. Methodology

3.1 Data Set

In our empirical analysis we use data on 6,059 firms, which are customers of the 8th largest Brazilian bank based on asset valuation. The dataset will provide the basis for the development and validation of the logistic regression models. The bank has credit operations with companies spread throughout the country. We employ data on firms for the period 2004 to 2005. Out of the bank’s 160,000 business customers approximately 24,000 are sufficiently large or important to have regular financial statements published according to Brazilian laws and financial reports which are carefully analysed by the risk department of the bank. These firms operate in various sectors of the economy but are subject to the same legal constraints for reporting financial information. More specifically, this means, that they use standard reports to publish their accounting information. Two types of information have been collected: (1) Financial reports such as Balance Sheets and Profit and Loss accounts; (2) Data derived from the client's account structure, and the client's credit standing as is known by the bank. This data was extracted for our sample of 6,059 firms, which were selected so as to include approximately 10% which defaulted during the period. In fact the actual number of defaulting companies included was 522.

3.2 Definition of default

Compared to previous studies, which use the criterion of bankruptcy, the definition of default employed in this paper is as described in Basel II. Therefore, default is considered to have occurred with regard to a particular obligor when one or more of the following events has taken place: (1) the obligor is more than 90 days overdue on any credit obligation; (2) the bank puts the credit obligation on non-accrued status; (3) the bank makes a charge-off or account-specific provision resulting from a significant perceived decline in credit worthiness subsequent to the bank taking on the exposure; (4) the bank consents to a distressed restructuring of the credit obligation; and (5) the obligor has sought or has been placed in bankruptcy or similar protection where this would avoid or delay repayment of the credit obligation.

3.3 Definition of Time horizon

The usual practice of the financial institutions is to establish credit limits for a period of one year following the recommendation of the 1988 Basel Accord. The same time criterion is used in this paper, as this requirement was not modified in the New Accord of 2003. Moreover, one year reflects the typical interval over which information on new obligors is requested by the banks in order to re-evaluate the credit limits. Furthermore, annual accounting statements are prepared to satisfy Brazilian authorities' requirements for corporate tax purposes and payments.

3.4 Variable Selection Process

In contrast to the majority of bankruptcy studies, the topic of variables selection is discussed and addressed more thoroughly in this paper. In particular, the selection process is not restricted to selection of financial ratios which have traditionally been employed in previous research. In addition to such conventional ratios, other financial ratios will be included according to a proposed framework that is devised later.

Essentially, what motivates us to reject the selection of traditional financial ratios solely is the crucial importance, at least in theory, of the selection of variables to the process of modelling. In our view, which is reinforced by the conclusions of Balcaen and Oghe (2004a), the selection of independent variables is an important challenge, which we believe will benefit from some theoretical ideas concerning company failure. In the next two subsections we present the arguments for the selected conventional and non-conventional ratios, respectively.

3.4.1 Selection of Conventional Financial Ratios

The selection of variables is often the most important part of modelling. Falkenstein *et al.* (2000) state that financial ratios are related to a firm's failure in the way that the speed of a car is related to the probability of crashing: there is a correlation, it's nonlinear, but there is no point at which failure is certain. Financial ratios have been intensively employed for modelling purposes during the last 40 years, mainly in bankruptcy prediction models. Usually, the ratios are grouped into various categories, but a great variety of classification is found in the literature. The most common classification divides the ratios into four sets: liquidity, profitability, gearing and activity ratios.

Despite the huge number of possible candidate ratios cited in the literature, there is no theoretical basis to dictate which predictors should be employed. All of the independent variables cannot be used, so one must find a way to find an appropriate subset. In this study, the 22 candidate financial ratios listed in table 1 were initially selected in order to cover the dimensions of liquidity, activity, financial structure, profitability and growth. See appendix for more detailed definitions.

[Insert Table 1 about here]

3.4.2 Proposed framework of complementary financial ratios

There is an old saying in the credit market: "Debts are paid off with cash and not with properties". It is this sentiment about the importance of 'cash' and the way that 'cash' is generated that drives the theory behind our framework of complementary financial ratios. Although the order of the trading operations may change, generally a company's cash flow cycle can be described as follows: (1) first, goods are purchased; (2) from this point starts the inventory period (raw material, work in progress, finished goods) and simultaneously its expenses are incurred (cash out); (3) goods are sold; (4) suppliers are paid off (cash out); (5) customers monies due are collected (cash in); and (6) expenses with taxes over labour and sales are incurred (cash out). Obviously, companies may well have other operations in addition to trading which represent cash in and cash out

Unfortunately cash information is only very crudely represented in many published financial statements. In the case of Brazilian published financial reports, cash flow is only roughly defined as net profit plus depreciation and amortization. Without entering into a debate over the purposes for which this definition was created, we note that this definition does not allow for the dynamic complexities of the true cash flow.

Our aim is to seek better indicators of company's solvency that can be extracted using financial reports published according to the Brazilian rules. To do this we focus on the classification of current assets and liabilities, exploring exclusively aspects relating to the liquidity of the companies.

Heath (1980) argued that accounting practices that have developed under conditions existing at one point in time may become so firmly embedded in our thought processes that they will come to be regarded as natural or inevitable. He continues by saying that the practice of classifying assets and liabilities as current or non current began early in the last century, in response to the perceived needs of commercial bankers. He considers that this practice is a vestige of a bygone era and should be abandoned because it is misleading.

In theory, the typical classification of assets and liabilities as current and non current indicates a relative measure of the firm's potential to pay its debts as they come due. The higher the current ratio, the more liquid the firm is. However, a firm can experience financial distress and still have a robust current ratio since this measure is insensitive to the timing of cash movements – those received and paid out. Also, we cannot overlook the fact that an increasing current ratio may hide slow-moving inventories and sluggish collections.

We are aware that the current time-honoured classification was originally created solely for accounting purposes. However, there is no doubt that establishing appropriate classifications is a process connected with specific demands. Therefore, without further exploring the fundamentals of accounting principles and conventions - which is beyond the aim of this paper, we will propose a new framework based on the context outlined above. Our target is to find new indicators of solvency in response to the demands of this research rather than to search for true definitions.

To develop the proposed approach, raw data from the bank in-house databases were used and two kinds of balance sheet were elaborated. The first of them is arranged in a standard Brazilian format of publication and is summarised in Table 2. In the second format, summarised in Table 3, the same data is rearranged to reflect the cash flow ideas outlined earlier.

[Insert Tables 2 & 3 about here]

In this new proposed format the current assets and liabilities are each separated into two sub-groups in such a way that they fit more naturally according to the company's activities. In each case, the first sub-group relates to its financial operations and the second comes from the trading operations. In the first sub-group all sorts of short term financial assets and liabilities are classified: cash, deposit accounts, securities, bank loans, trade finance and related parties. In the second sub-group are the short term accounts connected with the main trading activity of the firm such as customers, inventories, prepayments, provision for doubtful debtors, suppliers, accrued expenses and taxes based on payroll and trading matters.

Following Heath's (1980) argument, liabilities should be disposed of on the basis of different types of credit sources - spontaneous and negotiated. Spontaneous sources are those which grow out of normal patterns of profitable operation. Negotiated sources are those requiring conscious effort or specific negotiation by owners or managers. This differentiation is very important for our purposes because they provide strong insights into the company management and its financial flexibility and ability to transform assets quickly into cash.

Hence we use the proposed new sub-groupings (financial short term assets (FSTA); trading operation assets (TOA); onerous short term liabilities (OSTL) and trading operation liabilities (TOL) to form two new measures to appear in our proposed new financial ratios, namely:

FINANCIAL OR ONEROUS WORKING CAPITAL (F/OWK) = (FSTA – OSTL)

NEEDS OF WORKING CAPITAL (NWK) = (TOA – TOL).

The third measure in our proposed framework is owner's working capital (OWK), defined as:

OWNER'S WORKING CAPITAL (OWK) = (SHAREHOLDER'S FUNDS + DEFERRED INCOME) – FIXED ASSETS.

The value of owner's working capital can be seen as the amount of shareholder's funds held by the company in its current assets. In other words, it represents the owner's resources invested in the operational cycle of the company. It is proposed as a potentially important variant on the classical measure 'working capital or net working capital' (current assets – current liabilities) which are incorporated into the traditional financial ratios, i.e. current ratio (CURR) and working capital/total assets (WKASS).

Finally we choose net sales (NET SALES) as they reflect more dynamic characteristics than, for example, total assets which have widely been employed as a denominator in financial ratios. From the leading author's experience, the book value of assets must sometimes be treated with care in Brazil. On the one hand there are many companies whose assets are underestimated. On the other hand, over estimation in revaluation of assets is a strategy used by many Brazilian firms in an attempt to signal better economic performance.

From these four variables, we propose six complementary financial ratios, as detailed in Table 4.

[Insert Table 4 about here]

4. Model Choice

As noted earlier, logistic regression has been selected for our empirical analyses. This section provides a brief overview of the technique of logistic regression and its use within this paper.

4.1 Logistic Regression

Logistic regression is a modelling technique used extensively on data mining applications. The dependent variable in logistic regression is binary, and usually can take the value 1 with a probability *p*, or the value 0 with probability 1-*p*, where *p* depends on the characteristics (X) of the firm. In credit scoring Y is a binary response variable where Y=1 represents the event of a firm defaulting and Y=0 otherwise. In order to simplify the notation, denote $p(x) = P(Y = 1 / X = x)$ as the conditional probability of Y=1 given the covariate X=*x*.

For the sake of exposition, we first consider the univariate case where *P* (firm going to default/X) is not linearly related to the explanatory variable x, so the relationship is assumed to take the form of a logistic curve:

$$p(x) = \frac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}} \ ,$$

whose values are restricted to the range [0, 1]. Taking natural logarithms of this relationship then gives an alternative form of the logistic regression equation, referred to the logit of *p(x)*:

$$\log it\,[\mathrm{p(x)}] = \ln\left[\frac{p(x)}{1 - p(x)}\right] = \alpha + \beta x$$

where α is a constant, β is the predictor coefficient.

In logistic regression, $p(x)/(1 - p(x))$ is referred to as the 'odds' that a firm defaults, i.e. the chance that it defaults compared to the chance that it does not default. One important property of a logistic regression model $p(x) = \frac{e^{\alpha + \beta x}}{1 + e^{\alpha + \beta x}}$ is that e^{β} is the multiplicative factor by which the odds of a firm defaulting would increase for unit increase in the variable *x*. This multiplicative factor is referred to as the 'odds ratio'.

In this study the intention is to predict the probability that a firm will default based on a set of K covariates, $X_1,...,X_K$. , which implies a multivariate context. This model follows along the same lines as the univariate case. In the multivariate case the probability that a firm with characteristics $X_1 = x_1,...,X_K = x_K$ defaults is given by:

$$p(x_1, ..., x_K) = \frac{e^{\alpha + \beta_1 x_1 + ... + \beta_K x_K}}{1 + e^{\alpha + \beta_1 x_1 + ... + \beta_K x_K}}$$

The logit function is again used:

$$\log it[p(x_1,......,x_K)] = \alpha + \beta_1 x_1 + ... + \beta_K x_K$$

and the odds ratio e^{β_i} is the multiplicative factor by which the odds of a firm defaulting would increase for unit increase in the variable x_i for each variable x_i.

4.2 Type I and Type II Errors

The predictive performance of a logistic regression model can be assessed by looking at the classification table which demonstrates both the correct and incorrect classifications of the dichotomous dependent variable.

As described by Falkenstein et al (2000), when classification tools are used, default risk models can err in one of two ways. On the one hand, the model can indicate that granting credit to a firm is low risk when, in fact, it is high (Error type I). This type of error corresponds to the assignment of high credit quality rating to borrowers who nevertheless will default. The cost to the bank can be the loss of principal and interest. On the other hand, the model can assign a low credit quality rating to a firm when, in fact, the quality is high (Error type II). Potential losses resulting from this Type II error (commercial mistake) include the loss of return and origination fees when loans are either turned down or lost through non-competitive bidding. There other indirect costs involved with type II errors such as angry customers who may tend to terminate their relationship, in which case the loss is an implicit forgone gain. It would be desirable to minimise the weighted sum of costs caused by errors, although in many practical instances the appropriate costs are not available.

For the purposes of this analysis, error types I and II have been weighted equally and the results presented are obtained by fitting a logistic regression model using SAS (version 9.1), specifying the model selection criterion as ‘Validation Misclassification’ since this method selects the model that has the smallest misclassification rate for the validation data set. Very similar results were obtained using other criteria.

5. Results

5.1 Training and Validation data sets

It is usually recommended that when data sets are of reasonable size, the data is split into two. A random sample - typically referred to as the training sample -equivalent to 70% of the data was selected for model estimation, and the remaining 30% was used for model validation. Table 5 shows the overall default rate for the 6059 companies in the data set to be 8.6%, and the defaulting rates for the two samples are almost identical to the overall rate.

[Insert Table 5 about here]

5.2 Results of Logistic Regression model using only conventional financial ratios

The misclassification rates associated with the logistic regression model using only the conventional financial ratios are shown in Table 6. An important point to draw attention to is the generally lower misclassification rates in the validation sample compared with the training sample, confirming that over-fitting is not a problem in this analysis. Also, whilst it is not appropriate to compare studies simply on the basis of error rates, we do note in passing that these error rates compare favourably with the best of the models reported in Aziz and Dar (2004).

[Insert Table 6 about here]

Table 7 depicts the coefficients of the logistic model fitted and the associated odds ratios. The results highlight an important problem associated with using this type of model. The problem is referred to as the 'wrong sign' problem and commonly occurs in studies where many independent variables are included. More specifically when the relationships between two independent variables and the response variable are individually positive, it often happens in a multivariate model that one of the two coefficients will be negative, i.e. has the wrong sign. In this present case, we would have expected that both bank debt (BKDEB) and financial leverage (FINLEV) would be positively related to the probability of default. However these two independent variables are highly correlated (0.814), and whilst FINLEV has the expected sign, the weaker univariate predictor – BKDEB – has the wrong sign.

[Insert Table 7 about here]

The same reasoning can be applied to the variables working capital/assets (WKASS) and current ratio (CURR) which are also highly correlated (0.693) and in this case the latter's coefficient has wrong sign. Similarly, the coefficient for funds from operations/sales (FFOSA) has the wrong sign and its highest correlation is 0.411 with INTCOV, whose coefficient has the correct sign. In summary, the coefficients of BKDEB, CURR and FFOSA have the wrong sign and are weaker variables compared to FINLEV, WKASS and INTCOV, respectively. When counter-intuitive results of this sort are obtained, the pre-selection of the predictor variables is often recommended to remove variables to reduce these multicollinearity problems. However removal of variables that have previously contributed to low error rates will usually have the effect of increasing the error rates in any subsequent models.

5.3 Results of Logistic Regression model using conventional and new complementary financial ratios

The misclassification rates associated with the logistic regression model using the conventional financial ratios plus the new complementary financial ratios are shown in Table 8. We again note by comparing the training and validation sample results that over-fitting is not a problem. Comparing the results of Table 8 with those in Table 6, it is clear that the addition of the complementary variables has brought about a sizeable reduction in the already low misclassification rates.

[Insert Table 8 about here]

Table 9 shows the coefficients of the fitted logistic regression model and the associated odds ratios. In terms of the variables included in the new model we note that only two of the six explanatory variables from the previous model (see table 7) remain, FINLEV and INTCOV. The other four conventional variables (BKDEB, FFOSA, CURR and WKASS) have been replaced by just three of the complementary variables (F/OWKSA, OWKSA and NWKSA).

[Insert Table 9 about here]

As well as achieving greater predictive accuracy with less explanatory variables, the signs of the coefficients in this model are consistent with a priori expectations. As the probability of default is being modelled, companies which have the highest ratios of financial/onerous working capital/sales (F/OWKSA), owner's working capital/sales (OWKSA) and profit/interest (INTCOV) are less likely to default. Conversely, companies with the highest ratios of financial leverage (FINLEV) and needs of working capital (NWKSA) are more likely to default.

In theory the impact of financial leverage sometimes can be ambiguous. For instance, firms which are highly leveraged may be at risk of bankruptcy if they are unable to make payments on their debts. They may also be unable to find new lenders. However high financial leverage can also increase the shareholders' return on their investment and often provides tax advantages associated with borrowing. An in-depth discussion of the pros and cons of this is beyond of the aim of this paper.

The odds ratios can be used to give an indication of the relative impact of each of the five factors. For a unit increase in financial/onerous working capital/sales (F/OWKSA) while controlling for other factors in the model, the odds of bankruptcy decrease by 92%. Similarly, a unit increases in interest cover (INTCOV) and owner's working capital/sales (OWKSA) providing odds of bankruptcy declines by 16% and 64% respectively. On the other hand, unit increases in financial leverage (FINLEV) and needs of working capital (NWKSA) increase the odds of bankruptcy by 62% and 88%, respectively.

There are two important issues from the accounting and finance perspective that are embedded in the empirical results. First, the five variables selected demonstrate that issues such as structure, liquidity and profitability are important factors when evaluating the risk of default. However it is clear that the management of short-term resources is the most relevant factors in discriminating between healthy and non-healthy companies. In other words, firms which use more spontaneous resources are more successful than those that utilise more funds from banks loans. Second, without denying the importance of the traditional financial ratios of liquidity – quick and current ratio- they did not prove to be good predictors in this study. They were demonstrably less able to explain default than ratios developed by the rearrangement of the balance sheet described previously. These points can be related specifically to the context of the Brazilian economy. Brazil has one of the highest real interest rates in the world. It is therefore not a surprise that three of the variables included in the model are related to interest (INTCOV) or increased use of short onerous resources (FINLEV and F/OWKSA).

6. Discussion

First and foremost this paper has shown that the New Basel Accord requirement that banks must have a robust system in place to validate their models regarding the risks to which they are exposed is achievable using a statistical modelling approach. The Basel II definition of default was used in place of bankruptcy alone and the resulting model accuracy compares favourably with the best results achieved in studies reported elsewhere, see Aziz and Dar (2004). Our initial reaction to this result was surprise, as we suspected that the wider concept of default, including firms in financial distress which have not deteriorated financially as far as firms in bankruptcy, would be more difficult to detect and hence to predict. However this proved not to be the case, and we therefore surmise that the financial reports of the 6,059 ‘important’ firms included in the study were generally up-to-date, error-free and sensitive to financial deterioration as well as to bankruptcy.

Another implication of the New Basel Accord is that banks around the world will have to apply models based on their specific data, taking into account, in their entirety, the risks to which they are exposed. In this study, this has meant that the analysis has been conducted using a large good quality dataset with a meaningful number of companies in default and with a good knowledge of the Brazilian corporate lending context. For instance Brazil is usually dependent on foreign investments or erratic capital. This situation divides the credit market into periods of international or domestic turmoil, provoking significant volatility in the financial markets. These circumstances, combined with the large Brazilian public deficit, might lead to crowding-out effects, thus leaving companies highly indebted at serious liquidity risk, especially those dependent on short term loans for operations. The real interest rate in Brazil has been the highest in the world because the Federal Government is the principal debtor in the market, constantly rolling over its enormous debt.

As has been seen earlier, the results achieved, even without the inclusion of the complementary set of financial ratios, were statistically very good (although difficult to interpret), and were comparable with those achieved in many previously reported studies. However, this level of 'local' knowledge of the business culture and of the available data has meant that new indicators of solvency could be envisaged and hence new financial ratios were created. In particular this research has also proposed and developed a new framework, motivated by the ideas of Heath (1980). This framework has addressed concerns about the accounting classification of current assets and liabilities and its contribution in evaluating the risk and the performance of companies. In theory, the current ratio is an indication of a company's ability to meet short-term debt obligations. The higher the ratio, the more liquid the company is. However, this measure does not reflect cash flow well and can lead to a mistaken interpretation. For example an increasing current ratio may mask slow-moving inventories and sluggish collections of cash. In the proposed framework current assets and liabilities accounts are therefore split into two sub-groups each, in such a way that the accounts can fit more naturally according to the company's activities. From this proposed rearrangement a group of non-conventional financial ratios has been devised.

Comparison of the models with and without the complementary set of non-conventional financial ratios clearly demonstrates the benefits of the new framework. Three of the complementary financial ratios replace four traditional ratios, the misclassification rate drops from about 1.8% to 0.5%, and the model coefficients all have sensible and transparent interpretations. We believe that this success is at least in part because this group of new financial ratios provides a better fit than traditional ratios to the characteristics which distinguish healthy and non healthy companies in the Brazilian corporate lending context.

Finally, we attempt to infer from the variables selected something about the context in which the Brazilian companies operate, that is an economy with prohibitive interest rates for financing needs of working capital. In particular there is strong evidence that companies which rely more intensively on spontaneous credit instead of onerous credit have substantial competitive advantages and hence will be less likely to have solvency problems. Furthermore these companies tend to be more profitable since spontaneous credits are free of cost or less expensive than onerous credits.

7. Conclusions and further work

This paper has presented a statistical model to predict insolvency according to the Basel II definition of default using financial reports from customers of a major Brazilian bank. A new framework of financial ratios based on the reclassification of current assets and liabilities is proposed.

Our main conclusions are:

(a) The Basel II defaulting definition can be successfully predicted using statistical models.

(b) The quality of the models developed demonstrates the potential and importance of developing models based on the reality in which each bank operates.

(c) Complementary non-conventional financial ratios have proved to be very useful and informative when introduced alongside the traditional ratios.

(d) The variables selected in the model from the accounting point of view represent aspects relating to the management of short term sources and uses, covering interest and leverage.

(e) Companies which rely more intensively on spontaneous credit instead of onerous credit have substantial comparative advantages.

The school of thought surrounding Basel II is that banks should develop separate models for the obligor and the facility. The obligor model should predict the probability of default (PD) based on default definition and the facility model should predict the loss given default (LGD). This paper addresses only the first model (PD) based on cross-sectional data. This research is on-going and a more extended dataset which comprises information from the years 2000 to 2005 has been analysed. This data set will allow us to compare models using cross-sectional and panel data.

There is also potential benefit in models that incorporate economic factors or variables such as interest rates, exchange rates and performance of specific economic segments should be considered.

Finally, it seems very likely that the topic addressed in this paper will remain very important in banking and finance, given the need to assess risks in a systematic and validated fashion. There is no doubt that credit scoring combines advantageous characteristics: it is more robust, transparent, objective, clear, faster, uniform, reliable, impartial, self oriented and cheaper than other traditional methods. Moreover, credit scoring methods easily meet the rules laid down in the New Basel Accord.

REFERENCES

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

Aziz, M.A. and Dar, H.A. (2004) Predicting corporate financial distress: Whither do we stand? Department of Economics, Loughborough University.

Banking for International Settlements-BIS (2003) The New Basel Capital Accord. Basel, Switzerland: Committee on Banking Supervision.

Back, P. (2005) Explaining financial difficulties based on previous payment behaviour, management background variables and financial ratios *European Accounting Review,* 14(4), pp. 839-868.

Balcaen, S. and Ooghe, H (2004a) 35 years of studies on business failure: an overview of the classical statistical methodologies and their related problems, Department of Accountancy and Corporate Finance, Ghent University.

Balcaen, S. and Ooghe, H. (2004b) Alternative methodologies in studies on business failure: do they produce better results than the classic statistical methods? Department of Accountancy and Corporate Finance, Ghent University.

Beaver, W. (1967a) Financial ratios predictors of failure. Empirical research in accounting: selected studies, *Journal of Accounting Research, Sup.*(4), pp.71-111.

Becchetti, L. and Sierra, J. (2002) Bankruptcy risk and productive efficiency in manufacturing firms, *Journal of Banking and Finance,*27(11), pp. 2099-2120.

Caouette, J. B. Altman, E.I. and Narayanan, P (1998), Managing credit risk: the next great financial challenge, (New York: John Wiley & Sons, Inc).

Caruana, J. (2003) A vision of stability within the Basel Accord. *Financial Times.* A3 pg. 5. 30/10/2003.

Charitou, A., Neophytou, E., and Charalambous, C. (2004) Predicting corporate failure: Empirical evidence for the UK *European Accounting Review,* 13(3), pp 465-497.

Du, Y (2003) Predicting credit rating and credit rating changes: A new approach, *Queen's School of Business, Queen's University.*

Falkenstein, E., Boral, A. and Carty, L.V. (2000) Risk Calc TM for private companies, Moody's Default Model. *Moody's Investors Service.*

Fanning, K. and Cogger, K. (1994) A comparative analysis of artificial neural networks using financial distress prediction. *Intelligent Systems in Accounting, Finance, and Management,* 3, 241-252.

Ferguson, R.D. (2001) Remarks on Credit Risk. *Annual Credit and Risk Conference*. New York.

Heath, L.C., (1980) Is really working capital really working? *Journal of Accountancy*.

Hekanaho, J. Back, B. Sere, K., Laitinien, T. (1998) Analysing Bankruptcy with Multiple Methods", American Association for Artificial Intelligence.

Hosmer, D. and Lemeshow, S., (1989) Applied Logistic Regression, (New York: *John Wiley & Sons).*

Keasey, K and Watson, R (1986) The prediction of small company failure: some behavioural evidence for the UK *Accounting and Business Research,* 17(65), pp. 49-57.

Laitinen, T and Kankaanpaa, M (1999) Comparative analysis of failure prediction methods: The Finnish case *European Accounting Review,* 8(1), pp.67-92.

Lehmann, B. (2003) Is it worth the while? The relevance of qualitative information in credit rating *Working paper presented at the EFMA 2003 Meetings*, Helsinki, pp. 1-25.

Lennox, C. (1999) Identifying failing companies: a re-evaluation of the logit, probit and DA approaches, *Journal of Economics and Business*, 51(4), pp. 347-364.

Lo, A. W., (1986) Logit versus discriminant analysis: a specification test and application to corporate bankruptcies, *Journal of Econometrics,* 31, pp.151-178.

Lussier, R.N. (1995) A non financial business success versus failure prediction model for young firms *Journal of Small Business Management* 33(1), pp.8-20.

McQuown, J.A. (2003) Managing corporate credit risk: catalytic effects of Basel II." *The RMA Journal, October*.

Ohlson, J. (1980) Financial ratios and the probabilistic prediction of bankruptcy, *Journal of Accounting Research,* 18(1), pp. 109-131.

O'Leary, D.E., (1998) Using neural networks to predict corporate failure, *International Journal of Intelligent Systems in Accounting, Finance and Management*, 7(3), pp. 187-197.

Schumway, (2001) T Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *The Journal of Business,* 74(1), pp. 101-124

Sheppard, J.P. (1994) Strategy and bankruptcy: an exploration into organizational death *Journal of Management*. 20(4), pp. 795-833.

Sobehart, J., and Stein, R. (2000b) Moody's Public Firm Risk Model: A Hybrid Approach to Modeling Short Term Default Risk, Moody's Investors Service.

Stein, R.M., (2002) Benchmarking Default Prediction Models: Pitfalls and Remedies in Model Validation**,** *Moody´s KMV.*

Stone, M. and Rasp, J., (1993) The assessment of predictive accuracy and model over-fitting, *Journal of Business Finance and Accounting*, 20(1), pp.125-131.

Tucker, J. (1996) Neural Networks versus logistic regression in financial modeling: a methodological comparison. *In Proceedings of the 1996 World First Online Workshop on Soft Computing, Nagoya University,* Japan, pp. 19-30.

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

Zavgren, C. (1985) Assessing the vulnerability to failure of American industrial firms: a logistic analysis, *Journal of Business Finance and Accounting*.

Zmijewski, M. E. (1984) Methodological issues related to the estimation of financial distress prediction models, *Journal of Accounting Research, Supplement on Current Econometric Issues in Accounting Research,* 59-82.

ACKNOWLEGMENTS

We thank the Banco do Estado do Rio Grande do Sul (Banrisul), especially its President, for authorising the use of data from the bank and the Universidade do Rio dos Sinos (Unisinos) for its contribution to financial support in this research.

Appendix 1

VARIABLES FORMULAE

01) (CURRENT ASSETS - STOCKS)/CURRENT LIABILITIES
02) CURRENT ASSETS/CURRENT LIABILITIES
03) (CURRENT ASSETS + LONG TERM RECEIVABLES)/(CURRENT LIABILITIES + LONG TERM LIABLITIES)
04) CURRENT LIABILITIES/SHAREHOLDER'S FUNDS
05) (CURRENT LIABILITIES + LONG TERM LIABILITIES)/SHAREHOLDER'S FUNDS
06) FIXED ASSETS/SHAREHOLDER'S FUNDS
07) BANK LOANS/SHAREHOLDER'S FUNDS
08) (STOCKS/COST OF GOOD SOLD) × 360
09) (CUSTOMERS/ SALES) × 360
10) (SUPPLIERS/PURCHASES) × 360
11) (08 + 09) = OPERATING CYCLE
12) (11 – 10) = FINANCIAL CYCLE
13) SALES/TOTAL ASSETS
14) PROFIT BEFORE INTEREST AND TAXES(PBIT)/SALES
15) ASSETS TURNOVER × NET PROFIT MARGIN = (13× 14)
16) FINANCIAL LEVERAGE × OPERATING RETURN ON EQUITY
17) NET PROFIT BEFORE INTEREST/INTEREST PAID
18) $(((SALES_1 / SALES_0)/INFLATION\ RATE))/-1 \times 100)$
19) (CURRENT LIABILITIES + LONG TERM LIABILITIES)/TOTAL ASSETS
20) (CURRENT ASSETS – CURRENT LIABILITIES)/TOTAL ASSETS
21) PROFIT BEFORE INTEREST AND TAXES/TOTAL ASSETS
22) (NET PROFIT + DEPRECIATION + AMORTIZATION)/NET SALES
23)((SHAREHOLDER'S FUNDS + DEFERRED INCOME) – FIXED ASSETS)/ NET SALES
24) (TRADING OPERATI0N ASSETS – TRADING OPERATIONS LIABILITIES)/NET SALES
25) (FINANCIAL SHORT TERM ASSETS – ONEROUS SHORT TERM LIABILITIES)/NET SALES

26) (TRADING OPERATI0N ASSETS – TRADING OPERATIONS LIABILITIES)/ ((SHAREHOLDER'S FUNDS + DEFERRED INCOME) – FIXED ASSETS)
27) (FINANCIAL SHORT TERM ASSETS – ONEROUS SHORT TERM LIABILITIES)/ ((SHAREHOLDER'S FUNDS + DEFERRED INCOME) – FIXED ASSETS)
28) (FINANCIAL SHORT TERM ASSETS – ONEROUS SHORT TERM LIABILITIES)/ (TRADING OPERATI0N ASSETS – TRADING OPERATIONS LIABILITIES)

Table 1: Conventional Financial Ratios

VARIABLES	MNEMONIC
01) QUICK RATIO	QUICKR
02) CURRENT RATIO	CURR
03) TOTAL LIQUID RATIO	TOTR
04) SHORT TERM DEBT RATIO	SHTDEB
05) FINANCIAL LEVERAGE	FINLEV
06) FIXED ASSET RATIO	FIXR
07) BANK DEBT RATIO	BKDEB
08) INVENTORY PERIOD (DAYS)	INVENT
09) CUSTOMER COLLECTION PERIOD (DAYS)	CUST
10) SUPPLIERS PERIOD (DAYS)	SUPPL
11) OPERATING CYCLE (DAYS)	OPERC
12) FINANCIAL CYCLE (DAYS)	FINC
13) ASSET TURNOVER	ASSTUR
14) NET PROFIT MARGIN	NPROFM
15) RETURN ON CAPITAL EMPLOYED	RTKEM
16) OPERATING RETURN ON EQUITY	OPRTEQ
17) INTEREST COVERAGE RATIO	INTCOV
18) SALES GROWTH	GROSA
19) TOTAL DEBT/TOTAL ASSETS	TDEBAS
20) WORKING CAPITAL/TOTAL ASSETS	WKASS
21) RETURN ON ASSETS	PBITAS
22) FUNDS FROM OPERATIONS/NET SALES	FFOSA

Table 2: Summary of a Standard Brazilian Balance Sheet

ASSETS	LIABILITIES
CURRENT ASSETS Cash and equivalents Customers Inventories Other	CURRENT LIABILITIES Bank loans Suppliers Payroll/Taxes/Contributions Other
LONG TERM RECEIVABLES	LONG TERM LIABILITIES
FIXED ASSETS Investments Property, plant and equipment Deferred charges	DEFERRED INCOME
	SHAREHOLDER’S FUNDS
TOTAL ASSETS	TOTAL LIABILITIES

Table 3: Summary of the proposed Balance Sheet

ASSETS	LIABILITIES
SHORT TERM ASSETS	SHORT TERM LIABILITES
FINANCIAL Cash and equivalents TRADING OPERATION Customers Inventories	ONEROUS Bank loans Related parties TRADING OPERATING Suppliers Payroll/taxes/contributions
LONG TERM RECEIVABLES	LONG TERM LIABILITIES
FIXED ASSETS Investments Property, plant and equipment Deferred charges	DEFERRED INCOME
	SHAREHOLDER’S FUNDS
TOTAL ASSETS	TOTAL LIABILITIES

Table 4: Set of Complementary Financial Ratios derived from the proposed framework

VARIABLES	MNEMONIC
23) OWNER'S WORKING CAPITAL/NET SALES	OWKSA
24) NEEDS OF WORKING CAPITAL/NET SALES	NWKSA
25) FINANCIAL OR ONEROUS WORKING CAPITAL/NET SALES	F/OWKSA
26) NEEDS OF WORKING CAPITAL/OWNER'S WORKING CAPITAL	NWKOWK
27) FINANCIAL OR ONEROUS WORKING CAPITAL/OWNER'S WORKING CAPITAL	F/OWKOWK
28) FINANCIAL OR ONEROUS WORKING CAPITAL/NEEDS OF WORKING CAPITAL	F/OWKNWK

Table 5: Training and Validation Data Sets

STATUS	TRAINING	%	VALIDATION	%	TOTAL	%
GOOD	3876	91.4	1661	91.3	5537	91.4
DEFAULT	364	8.6	158	8.7	522	8.6
TOTAL	4240	100.0	1819	100.0	6059	100.0

Table 6: Misclassification rates (%) using conventional ratios

SAMPLE	ERROR TYPE I	ERROR TYPE II	TOTAL ERROR
Training	12.09	0.83	1.79
Validation	12.66	0.72	1.76

Method: Stepwise; Link function: Logit; Criteria: Validation Misclassification

Table 7: Parameter estimates for logistic regression model using conventional ratios

Variables	Estimates (SE)	Wald	p-value	Odds ratio
Intercept	-4.55 (0.34)	174.70	<.0001	0.011
BKDEB	-0.96 (0.11)	77.04	<.0001	0.385
FFOSA	0.26 (0.10)	6.64	0.0100	1.303
FINLEV	1.37 (0.11)	158.02	<.0001	3.954
INTCOV	-0.25 (0.03)	94.04	<.0001	0.778
CURR	0.45 (0.14)	10.36	0.0013	1.562
WKASS	-2.14 (0.22)	92.20	<.0001	0.117

Table 8: Misclassification rates (%) using conventional + complementary financial ratios

SAMPLE	ERROR TYPE I	ERROR TYPE II	TOTAL ERROR
Training	3.57	0.26	0.54
Validation	2.53	0.30	0.49

Method: Stepwise; Link function: Logit; Criteria: Validation Misclassification

Table 9: Parameter estimates for logistic regression model using conventional + complementary financial ratios

Variables	Estimates (SE)	Wald	p-value	Odds ratio
Intercept	-5.76 (0.52)	120.50	<.0001	0.003
F/OWKSA	-2.53 (0.31)	64.46	<.0001	0.080
FINLEV	0.48 (0.10)	24.68	<.0001	1.621
INTCOV	-0.17 (0.03)	30.39	<.0001	0.842
OWKSA	-1.02 (0.17)	37.75	<.0001	0.360
NWKSA	0.63 (0.21)	9.18	0.0024	1.877