Twenty-five years of z-scores in the UK: do they really work?

Vineet Agarwal*

Cranfield School of Management

and

Richard J. Taffler

The Management School, University of Edinburgh

Version 5.1: December 7, 2005

*Corresponding author
Vineet Agarwal
Cranfield School of Management
Cranfield, Bedford
MK43 0AL

Tel: +44 (0) 1234 751122
Fax: +44 (0) 1243 752554
Email: vineet.agarwal@cranfield.ac.uk

The authors are indebted to their many colleagues and seminar and conference participants who have provided valuable comments on earlier drafts of this paper over several years.

Twenty-five years of z-scores in the UK: do they really work?

Abstract

Although copious statistical failure prediction models are described in the literature, appropriate tests of whether such methodologies really work in practice are lacking. Validation exercises conducted are, at best, only *ex post* with clear confusion demonstrated between true *ex ante* predictive ability and *ex post* sample classification. This paper describes the operating characteristics of a well known UK-based z-score model and evaluates its performance over the twenty-five year period since it was originally developed. The model is shown to have true *ex ante* predictive ability over this extended time period and dominates more naïve prediction approaches. However, its performance is attenuated in the most recent period. Prima facie, such results also demonstrate the predictive ability of the published accounting numbers used in the z-score model calculation.

Keywords: *bankruptcy prediction, financial ratios, type I and type II errors, temporal stability*

Twenty-five years of z-scores in the UK: do they really work?

1. Introduction

There is renewed interest in credit risk assessment methods following Basel II and recent high profile failures such as Enron and Worldcom. New approaches are continuously being proposed (e.g., Hillegeist et al., 2004; Vassalou and Xing, 2004; Bharath and Shumway, 2004) and academic journals publish special issues on the topic (e.g., Journal of Banking and Finance, 2001). The traditional z-score technique for "predicting" corporate financial distress, however, is still a well accepted tool for practical financial analysis. It is discussed in detail in most of the standard texts and continues to be widely used both in the academic literature and by practitioners.

The z-score is used as a basic research tool in exploring such areas as merger and divestment activity (e.g. Shrieves and Stevens, 1979; Lasfer et al., 1996; Sudarsanam and Lai, 2001), asset pricing and market efficiency (e.g. Altman and Brenner, 1981; Katz et al., 1985; Dichev, 1998; Griffin and Lemmon, 2002; Ferguson and Shockley, 2003), capital structure determination (e.g. Wald, 1999; Graham, 2000; Allayannis et al., 2003; Molina, 2005), the pricing of credit risk (see Kao, 2000 for an overview), distressed securities (e.g. Altman, 2002: ch. 22; Marchesini et al., 2004), and bond ratings and portfolios (e.g. Altman, 1993: ch. 10; Caouette et al., 1998: ch 19). Z-score models are also extensively used as a tool in assessing firm financial health in going-concern research (e.g. Citron and

Taffler, 1992; Carcello et al., 1995; Mutchler et al., 1997; Louwers, 1998; Citron and Taffler, 2001 and 2004; Taffler et al., 2004).

Interestingly, however, no study to our knowledge has properly sought to test the predictive ability of the z-score approach in the almost 40 years since Altman's (1968) seminal paper was published. The existing literature that seeks to do this, at best, typically uses samples of failed and non-failed firms (e.g. Begley et al., 1996), rather than testing the respective models on the underlying population. This, of course, does not provide a true test of *ex ante* forecasting ability as the key issue of type II error rates (predicting non-failed as failed) is not addressed.[1]

This paper seeks to fill this important gap in the literature by specifically exploring the question of whether a well-established and widely-used UK-based z-score model driven by historic accounting data has true *ex ante* predictive ability or only *ex post* failure classification power over the 25 years since it was originally developed.

The remainder of the paper is structured as follows. Section 2 provides a brief overview of conventional z-score methodology and describes the UK-based model originally developed in 1977 (see Taffler, 1983) which is the subject of our analysis. Section 3 asks what is meant by the term corporate failure in this context and examines the track record of the z-score model since it was developed. Section 4 discusses what such models can and cannot do and section 5 reviews the

[1] The only possible exception is the recent paper of Beaver et al. (2005) for US data. However, their out-of-sample testing is for a much shorter period than this study, and their focus is not on the predictive ability of published operational models.

issues relating to temporal stability of such models. The final section, section 6, provides some concluding remarks.

2. The z-score model

The generic z-score is the distillation into a single measure of a number of appropriately chosen financial ratios, weighted and added. If the derived z-score is above a calculated score, the firm is classified as financially healthy, if below the cut-off, it is typically viewed as a potential failure.

Figure 1 here

This multivariate approach to failure prediction was first published almost 40 years ago with the eponymous Altman (1968) z-score model in the US, and there is an enormous volume of studies applying related approaches to the analysis of corporate failure internationally.[2] This paper reviews the track record of a well known UK-based z-score model for analysing the financial health of firms listed on the London Stock Exchange which was originally developed in 1977; a full description is provided in *Accounting and Business Research* volume 15, no. 52 (Taffler, 1983). The model itself was originally developed to analyse industrial (manufacturing and construction) firms only with separate models developed for

[2] For example, Altman and Narayanan (1997) review 44 separate published studies relating to 22 countries outside the US.

retail and service enterprises.[3] However, we apply it across all non-financial firms in the performance tests below.[4]

The first stage in building this model was to compute over 80 carefully selected ratios from the accounts of samples of 46 failed and 46 solvent industrial firms. Then using, inter alia, stepwise linear discriminant analysis, the z-score model was derived by determining the best set of ratios which, when taken together and appropriately weighted, distinguished optimally between the two samples.[5]

If a z-score model is correctly developed, its component ratios typically reflect certain key dimensions of corporate solvency and performance.[6] The power of such a model results from the appropriate integration of these distinct dimensions weighted to form a single performance measure, using the principle of the whole being worth more than the sum of the parts.

Table 1 provides the Taffler (1983) model's coefficients and ratios. It also indicates the four key dimensions of the firm's financial profile the ratios selected by the methodology are measuring: profitability, working capital position, financial risk and liquidity, which are identified by factor analysis, and the relative contribution of each to the overall discriminant power of the model using the

---

[3] Taffler (1984) also describes a model for analysing retail firms. His unpublished service company model is similar in form.

[4] Altman's (1968) model was also originally developed from samples of industrial companies alone but has conventionally been applied across the whole spectrum of non-financial firms.

[5] Data was transformed and Winsorised and differential prior probabilities and misclassification costs were taken into account in deriving an appropriate cut-off between the two groups. The Lachenbruch (1967) hold-out test provided two apparent classification errors.

[6] Factor analysis of the underlying ratio data should be undertaken to ensure collinear ratios are not included in the model leading to lack of stability and sample bias, and to help interpret the resulting model component ratios.

Mosteller-Wallace criterion. Profitability accounts for around 50% of the discriminant power and the three balance sheet measures together a similar proportion.

Table 1 here

In the case of this model, if the computed z-score is positive, i.e. above the "solvency threshold" on the "solvency thermometer" of figure 1, the firm is solvent and is very unlikely indeed to fail within the next year. However, if its z-score is negative, it lies in the "at risk" region and the firm has a financial profile similar to previously failed businesses and, depending on how negative, a high probability of financial distress. This may take the form of administration (Railtrack and Mayflower), receivership (Energis), capital reconstruction (Marconi), rescue rights issue, major disposals or spin-offs to repay creditors (Invensys), government rescue (British Energy), or acquisition as an alternative to bankruptcy.

Various statistical conditions need to be met for valid application of the methodology.[7] In addition, alternative methodologies such as quadratic discriminant analysis (e.g. Altman et al., 1977), logit and probit models (e.g. Ohlson, 1980; Zmijewski, 1984; Zavgren, 1985), mixed logit (Jones and Hensher, 2004), recursive partitioning (e.g. Frydman et al., 1985), hazard models (Shumway, 2001; Beaver et al., 2005) and neural networks (e.g. Altman et al.,

[7] These are discussed in Taffler (1983) with regard to the model described here and more generally in Taffler (1982), Jones (1987) and Keasey and Watson (1991) and need not detain us here.

1994) are used. However, since the results generally do not differ from the conventional linear discriminant model approach in terms of accuracy, or may even be inferior (Hamer, 1983; Lo, 1986; Trigueiros and Taffler, 1996), and the classical linear discriminant approach is quite robust in practice (e.g. Bayne et al., 1983) associated methodological considerations are of little importance to users.[8]

3. Forecasting ability

Since the prime purpose of z-score models, implicitly or explicitly, is to forecast future events, the only valid test of their performance is to measure their true *ex ante* prediction ability.[9] This is rarely done and when it is, such models may be found lacking. This may be because significant numbers of firms fail without being so predicted (type I errors). However, more usually, the percentage of firms classified as potential failures that do not fail (type II errors) in the population calls the operational utility of the model into question.[10] In addition, statistical evidence is necessary that such models predict better than chance, straight classification of all cases as non-failed or other simple strategies (e.g.

8 Trigueiros and Taffler (1996) point out there is little real evidence that such artificial intelligence neural network approaches, despite the claims of their proponents, dominate conventional multivariate models in such well-structured decision tasks as corporate bankruptcy prediction, particularly in the case of out-of-sample prediction. There are also issues of interpretation and practical utility.

9 Whereas techniques such as the Lachenbruch (1967) jackknife method, which can be applied to the original data to test for search and sample bias, are often used, inference to performance on other data for a future time period cannot be made because of potential lack of population stationary.

10 For example, the Bank of England model (1982) was classifying over 53% of its 809 company sample as potential failures in 1982, soon after it was developed.

prior year losses). Testing models on the basis of how well they classify failed firms *ex post* is not the same as true *ex ante* prediction tests.[11]

3.1. What is failure?

A key issue, however, is what is meant by corporate failure. Demonstrably, administration, receivership or creditors' voluntary liquidation constitute insolvency.[12] However, there are alternative events which may approximate to, or are clear proxies for, such manifestations of outright failure and result in loss to creditors and/or shareholders. Capital reconstructions, involving loan write-downs and debt-equity swaps or equivalent, can equally be classed as symptoms of failure, as can be acquisition of a business as an alternative to bankruptcy or major closures or forced disposals of large parts of a firm to repay its bankers. Other symptoms of financial distress, more difficult to identify, may encompass informal government support or guarantees, bank intensive care monitoring or loan covenant renegotiation for solvency reasons, etc. Nonetheless, in the analysis in this paper, we work exclusively with firm insolvencies on the basis these are clean measures, despite likely weakening in the predictive ability of the z-score model.

[11] Good examples are Begley et al., (1996) who conduct out-of-sample tests of type I and type II error rates for 1980s failures for both the Altman (1968) and Ohlson (1980) models and Altman (2002: 17-18), who provides similar sample statistics for his 1968 model through to 1999. However, neither study allows the calculation of true *ex ante* predictivity ability as in this paper, the acid test of such model purpose.

[12] The term bankruptcy used in the US, applies only to persons in the UK.

3.2. Track record over time

To assess the z-score model's performance in practical application, z-scores for the full population of non-financial firms available electronically and fully listed on the London Stock Exchange for at least two years at any time from 1979 (subsequent to when the model was developed) and 2003, a period of 25 years, are computed.[13] During this period there were 227 failures in our sample; 214 of these firms (94.3%) had z-scores<0 based on their last published annual accounts prior to failure indicating they had potential failure profiles.[14]

3.3. The population risk profile

The above results, however, are misleading. We need to know the percentage of the population to which the z-score model is applied classified as at risk (z<0), i.e. with financial profiles more similar to the failed group of firms from which the model was developed. This is a necessary, but not a sufficient, condition for subsequent financial distress. Figure 2 shows the percentage of such firms in our sample which varies over time. The low of 14% is registered in 1979 and the graph peaks at 42% in 2002, higher than the peaks of 30% in 1993 and 27% in

---

13 The accounting data required for model ratio calculations was primarily collected from the Thomson Financial *Company Analysis* and *EXSTAT* financial databases which between them have almost complete coverage of UK publicly fully listed companies. For the small number of cases not covered, *MicroEXSTAT* and *Datastream* were also used in that order.

14 Of the 13 firms misclassified, 10 had negative z-scores on the basis of their latest available interim/preliminary accounts prior to failure. On this basis, only three companies could not have been picked up in advance, including Polly Peck, where there were serious problems with the published accounts. Among other issues, there is a question mark over a missing £160m of cash and even the interim results, published only 17 days before Polly Peck's shares were suspended, show profits before tax of £110m on turnover of £880m. Whereas, as argued below, such multivariate models are quite robust to window dressing, this obviously cannot apply to major fraud.

1983 at the depths of the two recessions. The overall average is 26%.

Figure 2 here

The annual population failure rate measured as the percentage of firms failing over the *next* 12 months is provided in table 2. This reaches a high in 2001 of 2.3%, compared with rates of 2.1% in 1991 and 1.7% in 1990, years of deep recession. The overall average annual rate is 0.9%.

Table 2 here

3.4. True ex ante predictive ability

Low type I errors, however, are not an adequate test of the power of such models. A statistical comparison needs to be made with simple alternative classification rules. Also, misclassification costs need to be properly taken into account. In addition, we need to consider if the magnitude of the negative z-score has further predictive content.

3.4.1 Comparison with proportional chance model

Only a proportion of such firms at risk, however, will suffer financial distress. Knowledge of the population base rate allows explicit tests of the true *ex ante* predictive ability of the model where the event of interest is failure *in the next year*. This is essentially a test of whether the model does better than a proportional chance model which randomly classifies all firms as failures or non-failures based on population failure rates.

Table 2 shows that an average of 9 firms failed each year, and 214 of the 227 had z<0 on the basis of their last full year accounts before failure. In total, over the 25 year period, there were 6,733 firm years with z<0 and 18,955 with z > 0. The table also shows the overall conditional probability of failure given a negative z-score to be 3.2%. This differs significantly to the base failure rate of 0.9% at better than $\alpha = 0.001$ (z = 20.1).[15] Similarly, the conditional probability of non-failure given a positive z-score is 99.9% which is significantly different to the base rate of 99.1% at better than $\alpha = 0.001$ (z = 12.0).[16] In addition, on a 2x2 contingency table basis, the computed χ^2 statistic is 548.6 and strongly rejects the null hypothesis of no association between failure and z-score. Thus, this z-score model possesses true forecasting ability on this basis.

3.4.2. Comparison with simple loss-based classification rule[17]

However, the proportional chance model is probably too naïve and the true utility of the z-score model needs to be compared to some simple accounting-based model. We therefore classify firms with negative profit before tax (PBT) as potential failures and those with PBT>0 as non-failures. Table 3 provides the results (comparable to table 2) of using this classification criterion. It shows less than two thirds of the 227 failures over the 25 year period registered negative PBT on the basis of their last accounts before failure. In total there were 3,831 firm

---

15 $z = (p-\pi)/\sqrt{\pi(1-\pi)/n}$ where p = sample proportion, π = probability of chance classification and n = sample size. For the conditional probability of failure given z<0, p = 0.0318, π = 0.0088 and n = firms with z<0 = 6,773.

16 For the conditional probability of non-failure given z>0 at the beginning of the year, p = 0.9993, π = 0.9912 and n = 18,955.

years with PBT<0 and 21,857 with PBT>0. On this basis, the overall conditional probability of failure given a negative PBT is 3.7%, which differs significantly to the base failure rate of 0.9% at better than $\alpha = 0.001$ ($z = 18.7$).[18] Similarly, the conditional probability of non-failure given a positive PBT is 99.6%, which differs significantly to the base rate of 99.1% at better than $\alpha = 0.001$ ($z = 7.8$).[19] The 2x2 contingency table χ^2 statistic is 409.7 and strongly rejects the null hypothesis of no association between failure and loss in the last year. On this basis, a simple PBT-based model also appears to have true forecasting ability. The contingency coefficient for the degree of association between last year profit and subsequent failure/non-failure is 0.125 and is little different to that for the z-score model (0.145). In fact, the overall correct classification rate of this simple model is 85.3% dominating the 74.6% rate for the more complicated z-score model.

Table 3 here

3.4.3. Differential misclassification costs

The overall correct classification rates, however, are of little use. For instance, characterising all firms as non-failed would have led to no less than a 99.1% accuracy rate. In the credit market, the costs of misclassifying a firm that fails (type I error) is not the same as the cost of misclassifying a firm that does not fail

---

17 The authors are indebted to Steven Young for this suggestion.

18 $z = (p - \pi) / \sqrt{\pi(1-\pi)/n}$ where p = sample proportion, π = probability of chance classification and n = sample size. For the conditional probability of failure given PBT<0, $p = 0.0371$, $\pi = 0.0088$ and n = firms with PBT<0 = 3,831.

19 For the conditional probability of non-failure given PBT>0 at the beginning of the year, $p = 0.9961$, $\pi = 0.9912$ and $n = 21,857$.

(type II error). In the first case, the lender can lose up to 100% of the loan amount while, in the latter case, the loss is just the opportunity cost of not lending to that firm.

In assessing the practical utility of failure prediction models' ability, then, differential misclassification costs need to be explicitly taken into account. We compare the expected total costs of the z-score model (z) to the PBT model, the proportional chance model (PC) and the naïve model (Naïve) that classifies all firms as non-failed.

The total expected costs (EC) of decision-making based on the four different models are thus:

$$EC_z = p_2 * t_{II} * c_{II} + p_1 * t_I * c_I$$

$$EC_{PBT} = p_2 * t_{II} * c_{II} + p_1 * t_I * c_I$$

$$EC_{PC} = p_1 * p_2 * c_{II} + p_1 * p_2 * c_I$$

$$EC_{Naive} = p_1 * c_{II}$$

where:

p_1 = probability of failure

$p_2 = (1 - p_1)$ = probability of non-failure

t_I = type I error rate

t_{II} = type II error rate

c_I = cost of type I error

c_{II} = cost of type II error

Table 4 presents the overall accuracy rates, type I and type II error rates and total expected costs of decision-making using each of the four models employing representative values of c_I:c_{II} and p_1 = 0.9%, the *ex post* average annual failure rate over the 25-year period. Figure 3 represents graphically the expected costs for the four models for different type I / type II ratio error costs. It shows that no model is universally best and the total expected cost depends upon the differential costs of type I and type II errors. In fact, if the cost of making a type I error is <26x the cost of making a type II error, the naïve model gives the lowest total expected cost, while the PBT model gives the lowest expected cost if the ratio is between 26x and 40x. The z-score model adds value to the decision-making process only if the ratio of c_I:$c_{II} \geq 40$.[20]

Table 4 and Figure 3 here

3.4.4. Differential misclassification costs, prior probabilities and cut-off point

The analysis in section 3.4.3 above is incomplete as changing the c_I:c_{II} ratio leads to changes in the z-score cut-off. The optimal cut-off point for a discriminant model (e.g. Altman et al., 1977) is given by:

$$z_c = \ln\left(\frac{p_1}{p_2} * \frac{c_I}{c_{II}}\right)$$

with p_1, p_2, c_I and c_{II} as defined previously.

[20] The proportional chance model is dominated by one of the other three models across all values of c_I:c_{II}.

Taking this into account and again setting p_1 equal to the average empirical failure rate over our 25-year period, table 5 presents the different cut-off points for different cost ratios. Now, total expected costs associated with the z-score model are always lower than using PBT<0 when the cut-off point is adjusted to reflect the different costs ratio.

Table 5 here

3.4.5. Probability of failure and severity of negative z-score

Most academic research in this field has focused exclusively on whether the derived z-score is above or below a particular cut-off. However, does the magnitude of the (negative) z-score provide further information on the actual degree of risk of failure within the next year for z<0 firms?

To explore whether the z-score construct is an ordinal or only a binary measure of bankruptcy risk, we explore failure outcome rates by negative z-score quintiles over our 25-year period. Table 6 provides the results.

Table 6 here

As can be seen, there is a monotononic relationship between severity of z-score and probability of failure in the next year which falls from 7.3% in the worst quintile of z-scores to 0.8% for the least negative quintile. Overall, the weakest 20% of negative z-scores accounts for 42% of all failures and the lowest two quintiles together capture over two thirds (68%) of all cases. A contingency table test of association between z-score quintile and failure rate is highly significant

(χ^2 = 105.9). As such, we have clear evidence the worse the negative z-score, the higher the probability of failure; the practical utility of the z-score is clearly significantly enhanced by taking into account its magnitude.

4. What z-score models can and cannot do

Z-scores, for some reason, appear to generate a lot of emotion and attempts to demonstrate they do not work (e.g. Morris, 1997). However, much of the concern felt about their use is based on a misunderstanding of what they are and are not and what they are designed to do and not do.

4 .1 What a z-score model is

Essentially, a z-score is descriptive in nature. It is made up of a number of fairly conventional financial ratios measuring important and distinct facets of a firm's financial profile, synthesised into a single index. The model is multivariate, as are a firm’s set of accounts, and is doing little more than reflecting and condensing the information they provide in a succinct and clear manner.

The z-score is primarily a readily interpretable communication device, using the principle that the whole is worth more than the sum of the parts. Its power comes from considering the different aspects of economic information in a firm’s set of accounts simultaneously, rather than one at a time, as with conventional ratio analysis. The technique quantifies the degree of corporate risk in an independent, unbiased and objective manner. This is something it is difficult to do using judgement alone.

4.2 What it is not

A negative z-score is, strictly speaking, *not* a prediction of failure and the z-score model should not be treated in practical usage as a prediction device. What the statistical model is asking is "does this firm have a financial profile more similar to the failed group of firms from which the model was developed or the solvent set?" A negative z-score is only a necessary condition for failure, not a sufficient one, as table 2 demonstrates.

4.3. Philosophical issues

Z-score models are also commonly censured for their perceived lack of theory. For example, Gambling (1985 :420) entertainingly complains that:

> "… this rather interesting work (z-scores) … provides no theory to explain insolvency. This means it provides no pathology of organizational disease…. Indeed, it is as if medical research came up with a conclusion that the cause of dying is death…. This profile of ratios is the corporate equivalent of… 'We'd better send for his people, sister', whether the symptoms arise from cancer of the liver or from gunshot wounds."

However, once again, critics are claiming more for the technique than it is designed to provide. Z-scores are not explanatory theories of failure (or success) but pattern recognition devices. The tool is akin to the medical thermometer in indicating the probable presence of disease and assisting in tracking the progress of and recovery from such organisational illness. Just as no one would claim this simple medical instrument constitutes a scientific theory of disease, so it is only misunderstanding of purpose that elevates the z-score from its simple role as a measurement device of financial risk, to the lofty heights of a full-blown theory of corporate financial distress.

Nonetheless, there are theoretical underpinnings to the z-score approach, although it is true more research is required in this area. For example, Scott (1981) develops a coherent theory of bankruptcy and, in particular, shows how the empirically determined formulation of the Altman et al. (1977) ZETA™ model and its constituent variable set fits the postulated theory quite well. He concludes (p341) "Bankruptcy prediction is both empirically feasible and theoretically explainable". Taffler (1983) also provides a theoretical explanation of the model described in this paper and its constituent variables drawing on the well established liquid asset (working capital) reservoir model of the firm which is supplied by inflows and drained by outflows. Failure is viewed in terms of exhaustion of the liquid asset reservoir which serves as a cushion against variations in the different flows. The component ratios of the model measure different facets of this "hydraulic" system.

There are also sound practical reasons why this multivariate technique works in practice. These relate to (i) the choice of financial ratios by the methodology which are less amenable to window dressing by virtue of their construction, (ii) the multivariate nature of the model capitalizing on the swings and roundabouts of double entry, so manipulation in one area of the accounts has a counterbalancing impact elsewhere in the model, and (iii) generally the empirical nature of its development. Essentially, potential insolvency is difficult to hide when such "holistic" statistical methods are applied.

5. Temporal stability

Mensah (1984) points out that users of accounting-based models need to recognise that such models may require redevelopment from time to time to take into account changes in the economic environment to which they are being applied. As such, their performance needs to be carefully monitored to ensure their continuing operational utility. In fact, when we apply the Altman (1968) model originally developed using firm data from 1945 to 1963 to non-financial US firms listed on NYSE, AMEX and NASDAQ between 1988 and 2003, we find almost half of these firms (47%) have a z-score less than Altman's optimal cut-off of 2.675. In addition, 19% of the firms entering Chapter 11 during this period had z-scores greater than 2.675.[21]

Nonetheless, it is interesting to note that, in practice, such models can be remarkably robust and continue to work effectively over many years, as convincingly demonstrated above. Altman (1993: 219-220) reports a 94% correct classification rate for his ZETA™ model for over 150 US industrial bankruptcies over the 17 year period to 1991, with 20% of his firm population estimated as then having ZETA™ scores below his cut-off of zero.

[21] Begley et al. (1996) report out-of-sample type I and type II error rates of 18.5% and 25.1% for the Altman (1968) model and 10.8% and 26.6% for the Ohlson (1980) model using small samples of bankrupt and non-bankrupt firms.

In the case of the UK-based z-score model reviewed in this paper, figure 2 shows how the percentage of firms with at-risk z-scores varies broadly in line with the state of the economy. However, it increases dramatically from around 25% in 1997 to 41% by 2003 rendering it a very blunt instrument over the last few years. Although the model is being applied to retail and service firms in addition to the industrial sector from which it was developed, nonetheless, table 2 does provide evidence that the z-score model is no longer working as well, not in terms, necessarily, of the percentage of type I errors, but in terms of the very high type II error rates.

Factors that might be driving this diminution in predictive power in the recent period include (i) the growth in the services sector associated with contraction in the number of industrial firms listed,[22] (ii) the doubling in the rate of loss-making firms between 1997 and 2002, as demonstrated in table 3, with one in four firms in 2002 making historic cost losses, even before amortisation of goodwill, (iii) increasing kurtosis in the model ratio distributions indicating reduced homogeneity in firm financial structures, and (iv) increasing use of new financial instruments (Beaver et al., 2005). There are also questions about the impact of significant changes in accounting standards and reporting practices over the life of the z-score model, although as the model is applied using accounting data on a standardised basis, any potential impact of such changes is reduced.

On this basis then, we have strong evidence that, despite the remarkably long track record of robust performance and predictive ability of the Taffler (1983) z-

[22] As indicated above, our z-score model was derived originally using exclusively industrial firm data.

score model demonstrated in our paper, it is now no longer satisfactory in application and needs to be redeveloped with recent data.

Begley et al. (1996), Hillegeist et al. (2004) and Bharath and Shumway (2004) recalculate Altman's original 1968 model updating his ratio coefficients on new data; however, they all find that their revised models perform less well than the original model. The key requirement is to redevelop such models using ratios that measure more appropriately the key dimensions of firms' current financial profiles reflecting the changed nature of their financial structures, performance measures and accounting regimes.

Our results, interestingly, are in contrast to Beaver et al. (2005) who claim their derived three variable financial ratio predictive model is robust over a 40-year period. However, their only true *ex ante* tests are based on a hazard model fitted to data from 1962 to 1993 which is tested on data over the following 8 years. As such, the authors are only in a position to argue for short-term predictive ability. Interestingly, their hazard model also performs significantly less well than the z-score model reported in this paper over a full 25-year out-of-sample time horizon.[23]

6. Conclusions

This study describes a widely-used UK-based z-score model and explores its track record over the twenty five year period since it was developed. It is the first study to conduct valid tests of the true predictive ability of such models explicitly.

[23] Beaver et al. (2005) also do not take into account differential misclassification costs.

The paper demonstrates that the z-score model described, which was developed in 1977, has had true failure prediction ability over at least a 20-year period subsequent to its development, and dominates alternative, simpler approaches. Such techniques, if carefully developed and tested, continue to have significant value for financial statement users concerned about corporate credit risk and firm financial health. They also demonstrate the predictive ability of the underlying accounting data when correctly read in an holistic way. The value of adopting a formal multivariate approach, in contrast to ad hoc conventional one-at-a-time financial ratio calculation in financial analysis, is evident. Nonetheless, this paper does confirm that such failure prediction models need to be re-developed periodically to maintain operational utility.

References

Allayannis, G., Brown, G.W. and Klapper, L.F. (2003). 'Capital structure and financial risk: evidence from foreign debt use in East Asia'. *Journal of Finance,* 58(6):2667-2709.

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

____________. (1993). *Corporate financial distress and bankruptcy.* New York: John Wiley, 2nd edition.

____________. (2002). *Bankruptcy, credit risk, and high yield junk bonds: a compendium of writings.* Oxford: Blackwell Publishing.

____________, Haldeman, R.G. and Narayanan, P. (1977). 'ZETA analysis: A new model to identify bankruptcy risk of corporations'. *Journal of Banking and Finance*, 1(1):29-51.

Altman E.I. and Brenner, M. (1981). 'Information effects and stock market response to signs of firm deterioration'. *Journal of Financial and Quantitative Analysis,* 16(1):35-52.

Altman E.I., Marco, G. and Varetto, F. (1994). 'Corporate distress diagnosis: comparisons using linear discriminant analysis and neural networks (the Italian experience)'. *Journal of Banking and Finance*, 18(3):505-529.

Altman E.I. and Narayanan, P. (1997). 'An international survey of business failure classification models'. *Financial Markets and Institutions,* 6(2):1-57.

Bank of England. (1982). 'Techniques for assessing corporate financial strength'. *Bank of England Quarterly Bulletin*, June:221-223.

Bayne, C.K., Beauchamp, J.J., Kane, V.E. and McCabe, G.P. (1983). 'Assessment of Fisher and logistic linear and quadratic discriminant models'. *Computational Statistics and Data Analysis,* 1:257-273.

Beaver, W.H., McNichols, M.F. and Rhie, J. (2005). 'Have financial statements become less informative? Evidence from the ability of financial ratios to predict bankruptcy'. *Review of Accounting Studies*, 10(1):93-122.

Begley, J., Ming, J. and Watts, S. (1996). 'Bankruptcy classification errors in the 1980s: an empirical analysis of Altman's and Ohlson's models'. *Review of Accounting Studies*, 1(4):267-284.

Bharath, S.T. and Shumway, T. (2004). 'Forecasting Default with the KMV-Merton Model'. Working paper, University of Michigan.

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

Carcello, J.V., Hermanson, D.R. and Huss, H.F. (1995). 'Temporal changes in bankruptcy-related reporting'. *Auditing: A Journal of Practice and Theory*, 14(2):133-143.

Citron, D.B. and Taffler, R.J. (1992). 'The audit report under going concern uncertainties: an empirical analysis'. *Accounting and Business Research,* 22(88):337-345.

________________________. 'Ethical behaviour in the U.K. audit profession: the case of the self-fulfilling prophecy under going-concern uncertainties'. *Journal of Business Ethics,* 29(4):353-363.

________________________. (2004). 'The comparative impact of an audit report standard and an audit going-concern standard on going-concern disclosure rates'. *Auditing: A Journal of Practice and Theory,* 23(2):119-130.

Dichev, I.D. (1998). 'Is the risk of bankruptcy a systematic risk?' *Journal of Finance,* 53(3):1131-1147.

Ferguson, M.F. and Shockley R.L. (2003). 'Equilibrium "anomalies"'. *Journal of Finance*, 58(6):2549-2580.

Frydman, H., Altman, E.I. and Kao, D.G. (1985). 'Introducing recursive partitioning for financial classification: the case of financial distress'. *Journal of Finance,* 40(1):269-291.

Gambling, T. (1985). 'The accountant's guide to the galaxy including the profession at the end of the universe'. *Accounting Organisations and Society,* 10(4):415-426.

Graham, J.R. (2000). 'How big are the tax benefits of debt?' *Journal of Finance,* 55(5):1901-1941.

Griffin, J. and Lemmon, L. (2002). 'Book-to-market equity, distress risk, and stock returns'. *Journal of Finance*, 57(5):2317-2336.

Hamer, M. (1983). 'Failure prediction: sensitivity of classification accuracy to alternative statistical methods and variable sets'. *Journal of Accounting and Public Policy,* 2(4):289-307.

Hillegeist, S.A., Keating, E.K., Cram, D.P. and Lundstedt, K.G. (2004). 'Assessing the probability of bankruptcy'. *Review of Accounting Studies*, 9(1):5–34.

Jones, F.L. (1987). 'Current techniques in bankruptcy prediction'. *Journal of*

Accounting Literature, 6:131-164.

Jones, S. and Hensher, D.A. (2004). 'Predicting firm financial distress: a mixed logit model'. *The Accounting Review*, 79(4):1011-1038.

Journal of Banking and Finance (2001). Special issue on credit ratings and the proposed new BIS guidelines on capital adequacy for bank credit assets. 25(1).

Kao, D.L. (2000). 'Estimating and pricing credit risk: an overview'. *Financial Analysts Journal,* 56(4):50-66.

Katz, S., Lilien, S. and Nelson, B. (1985). 'Stock market behaviour around bankruptcy model distress and recovery predictions'. *Financial Analysts Journal,* 41(1):70-74.

Keasey, K. and Watson, R. (1991). 'Financial distress prediction models: a review of their usefulness'. *British Journal of Management,* 2(2):89-102.

Lachenbruch, P.A. (1967). 'An almost unbiased method of obtaining confidence intervals for the probability of misclassification in discriminant analysis'. *Biometrics,* 23(4):639-645.

Lasfer, M., Sudarsanam, P.S. and Taffler, R.J. (1996). 'Financial distress, asset sales, and lender monitoring'. *Financial Management,* 25(3):57-66.

Lo, A.W. (1986). 'Logit versus discriminant analysis'. *Journal of Econometrics,* 31(2):151-178.

Louwers, T.J. (1998). 'The relation between going-concern opinions and the auditor's loss function'. *Journal of Accounting Research*, 36(1):143-156.

Marchesini, R., Perdue, G. and Bryan, V. (2004). 'Applying bankruptcy prediction models to distressed high yield bond issues'. *Journal of Fixed*

Income, 13(4):50-56.

Mensah, Y. (1984). 'An examination of the stationarity of multivariate bankruptcy prediction models: a methodological study'. *Journal of Accounting Research,* 22(1):380-395.

Molina, C.A. (2005). 'Are firms underleveraged? An examination of the effect of leverage on default probabilities'. *Journal of Finance,* 60(3):1427-1459.

Morris, R. (1997). *Early warning indicators of corporate failure*. Ashgate Publishing.

Mutchler, J.P., Hopwood, W. and McKeown, J.C. (1997). 'The influence of contrary information and mitigating factors on audit opinion decisions on bankrupt companies'. *Journal of Accounting Research*, 35(2):295-301.

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

Scott, J. (1981). 'The probability of bankruptcy: a comparison of empirical predictions and theoretical models'. *Journal of Banking and Finance*, 5:317-334.

Shrieves, R.E. and Stevens D .L. (1979). 'Bankruptcy avoidance as a motive for merger'. *Journal of Financial and Quantitative Analysis,* 14(3):501-515.

Shumway, T. (2001). 'Forecasting bankruptcy more accurately: a simple hazard model'. *Journal of Business,* 74(1):101-124.

Sudarsanam, S. and Lai, J. (2001). 'Corporate financial distress and turnaround strategies: an empirical analysis'. *British Journal of Management,* 12(3):183-199.

Taffler, R.J. (1983). 'The assessment of company solvency and performance using a statistical model'. *Accounting and Business Research,* 15(52):295-308.

_________. (1984). 'Empirical models for the monitoring of UK corporations'. *Journal of Banking and Finance,* 8(2):199-227.

_________, Lu, J. and Kausar, A. (2004). 'In denial? Stock market underreaction to going-concern audit report disclosures'. *Journal of Accounting & Economics,* 38(1-3):263-296.

Trigueiros, D. and Taffler, R.J. (1996). 'Neural networks and empirical research in accounting'. *Accounting and Business Research,* 26(4):347-355.

Vassalou, M. and Xing, Y. (2004). 'Default risk in equity returns'. *Journal of Finance*, 59 (2):831-868.

Wald, J.K. (1999). 'How firm characteristics affect capital structure: an international comparison'. *Journal of Financial Research,* 22(2):161-187.

Zavgren, C.V. (1985). 'Assessing the vulnerability to failure of American industrial firms: a logistic analysis'. *Journal of Business Finance and Accounting,* 12(1):19-45.

Zmijewski, M.E. (1984). 'Methodological issues related to the estimation of financial distress prediction models'. *Journal of Accounting Research,* 22(Supplement):59-82.

Figure 1: The Solvency Thermometer

Firms with computed z-score < 0 are at risk of failure; those with z-score > 0 are financially solvent.

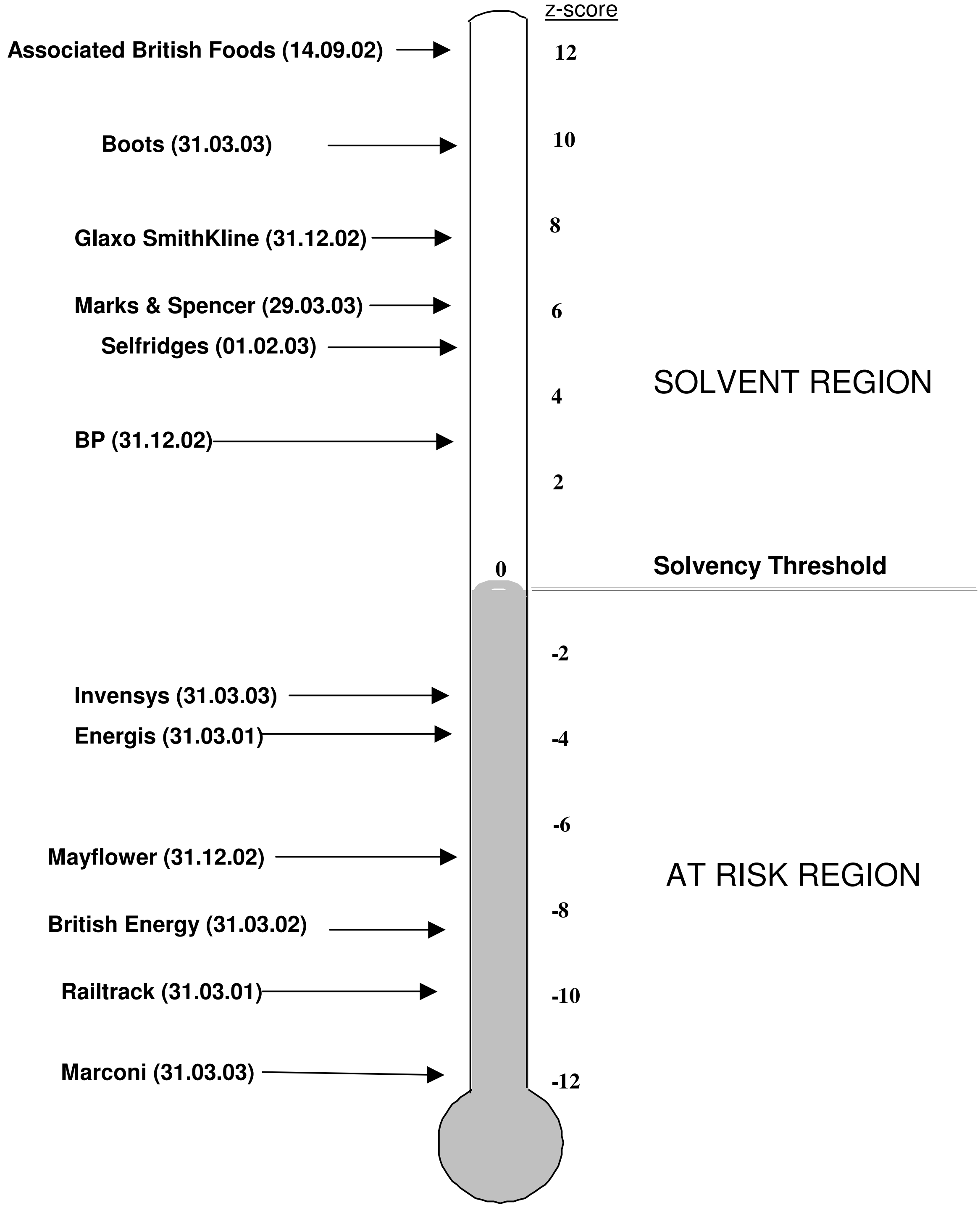

Figure 2: Percentage of firms at risk

The z-scores of all the firms in our sample are computed based on their last available full year accounts as at the end of September of each year from 1979 to 2003.

Figure 3: Expected costs of using different models

Z-scores and profit before tax (PBT) figures for all the firms in our sample are computed based on their last available full year accounts as at the end of September of each year from 1979 to 2003 (year t). Firms are then tracked for the next twelve months (to 30 September of year t+1) to identify those that failed. The z-score model classifies all firms with z<0 as potential failures, the PBT model classifies all firms with PBT<0 as potential failures, the proportional chance model randomly classifies firms as potentially failed/non-failed based on ex post determined probability of failure and the naïve model classifies all firms as non-failures. The type I error rate represents the percentage of failed firms classified as non-failed by the respective model, and the type II error rate represents the percentage of non-failed firms classified as failed by the respective model. Overall accuracy gives the percentage of firms correctly classified in total; c_I:c_{II} is the ratio of the relative costs of type I to type II errors. Total expected costs are based on the average type I and type II error rates and prior probability of failure based on the average failure rate over the 25-year period. For illustrative purposes, we assume the cost of a type II error (c_{II}) is 1%. Assuming a constant cost ratio (c_I:c_{II}), change in the type II error cost produces a proportional change in total expected costs.

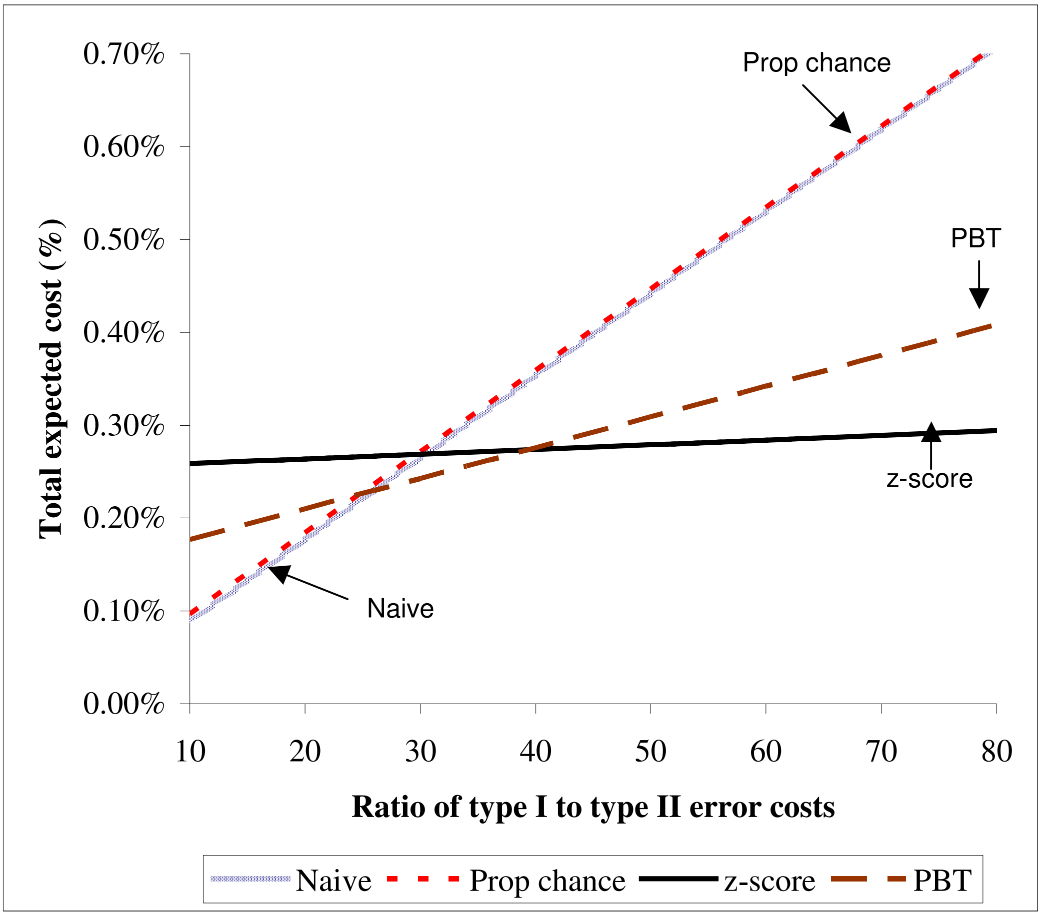

Table 1: Model for analysing fully listed industrial firms

The model takes the form:

$$z = 3.20 + 12.18 * x_1 + 2.50 * x_2 - 10.68 * x_3 + 0.029 * x_4$$

where

x_1 = profit before tax/current liabilities (53%)

x_2 = current assets/total liabilities (13%)

x_3 = current liabilities/total assets (18%)

x_4 = no-credit interval[1] (16%)

and $c_0 \ldots c_4$ are the respective model constant and coefficients. The percentages in brackets represent the Mosteller-Wallace contributions of the ratios to the power of the model. x_1 measures profitability, x_2 working capital position, x_3 financial risk, and x_4 liquidity.

[1] no-credit interval = (quick assets – current liabilities)/daily operating expenses with the denominator proxied by (sales – PBT – depreciation)/365

Table 2: Failure rates and percentage of firms with z<0

The z-scores of all the firms in our sample are computed based on their last available full year accounts as at the end of September of each year from 1979 to 2003 (year t). Firms are then tracked over the next twelve months (to 30 September of year t+1) to identify those that failed. Columns 2 and 5 give the number of firms with z<0 and z>0 respectively on 30 September of each year t, columns 3 and 6 provide the number of firms failing with -ve and +ve z-scores respectively between 1 October of year t and 30 September of year t+1. Column 4 gives the percentage of -ve z-score firms that failed and column 7 gives the percentage of +ve z-score firms that failed. Column 8 indicates the percentage of firms with -ve z-score on 30 September of year t and the last column gives the percentage of firms that failed between 1 October of year t and 30 September of year t + 1.

	z<0			**z>0**				**Overall**
Year t	**No. of firms**	**No. of failures**	**Failure rate (%)**	**No. of firms**	**No. of failures**	**Failure rate (%)**	**z<0 (%)**	**failure rate (%)**
1979	186	11	5.9	1157	0	0.0	13.8	0.8
1980	225	14	6.2	1095	2	0.2	17.0	1.2
1981	258	17	6.6	1014	0	0.0	20.3	1.3
1982	312	8	2.6	912	0	0.0	25.5	0.7
1983	325	14	4.3	864	0	0.0	27.3	1.2
1984	291	10	3.4	844	0	0.0	25.6	0.9
1985	258	4	1.6	827	0	0.0	23.8	0.4
1986	246	3	1.2	764	0	0.0	24.4	0.3
1987	211	2	0.9	742	0	0.0	22.1	0.2
1988	165	2	1.2	773	0	0.0	17.6	0.2
1989	191	11	5.8	762	1	0.1	20.0	1.3
1990	220	15	6.8	731	1	0.1	23.1	1.7
1991	274	20	7.3	659	0	0.0	29.4	2.1
1992	276	6	2.2	608	0	0.0	31.2	0.7
1993	305	5	1.6	638	0	0.0	32.3	0.5
1994	253	6	2.4	683	0	0.0	27.0	0.6
1995	249	6	2.4	719	0	0.0	25.7	0.6
1996	274	9	3.3	775	0	0.0	26.1	0.9
1997	282	10	3.5	822	0	0.0	25.5	0.9
1998	293	5	1.7	803	1	0.1	26.7	0.5
1999	314	8	2.5	706	0	0.0	30.8	0.8
2000	345	7	2.0	601	1	0.2	36.5	0.8
2001	333	16	4.8	529	4	0.8	38.6	2.3
2002	345	4	1.2	485	3	0.6	41.6	0.8
2003	302	1	0.3	442	0	0.0	40.6	0.1
Total	6733	214	3.2	18955	13	0.1	26.2	0.9

Table 3: Failure rates and percentages of loss-making firms

Profit before tax (PBT) figures for all firms in our sample are computed based on their last available full year profit and loss account as at the end of September of each year from 1979 to 2003 (year t). Firms are then tracked over the next twelve months (to 30 September of year t+1) to identify those that failed. Columns 2 and 5 give the number of firms with PBT<0 and PBT>0 respectively on 30 September of each year t; columns 3 and 6 provide the number of firms failing with -ve and +ve PBT respectively between 1 October of year t and 30 September of year t+1. Column 4 provides the percentage of -ve PBT firms that failed and column 7 the percentage of +ve PBT firms that failed. Column 8 indicates the percentage of firms with -ve PBT on 30 September of year t and the last column gives the percentage of firms that failed between 1 October of year t and 30 September of year t + 1.

	PBT<0			PBT>0				Overall
Year t	No. of firms	No. of failures	Failure rate (%)	No. of firms	No. of failures	Failure rate (%)	PBT<0 (%)	failure rate (%)
1979	134	7	5.2	1209	4	0.3	10.0	0.8
1980	152	4	2.6	1168	12	1.0	11.5	1.2
1981	235	13	5.5	1037	4	0.4	18.5	1.3
1982	263	5	1.9	961	3	0.3	21.5	0.7
1983	255	9	3.5	934	5	0.5	21.4	1.2
1984	147	8	5.4	988	2	0.2	13.0	0.9
1985	115	2	1.7	970	2	0.2	10.6	0.4
1986	118	1	0.8	892	2	0.2	11.7	0.3
1987	94	2	2.1	859	0	0.0	9.9	0.2
1988	66	2	3.0	872	0	0.0	7.0	0.2
1989	50	2	4.0	903	10	1.1	5.2	1.3
1990	80	4	5.0	871	12	1.4	8.4	1.7
1991	142	14	9.9	791	6	0.8	15.2	2.1
1992	180	4	2.2	704	2	0.3	20.4	0.7
1993	211	5	2.4	732	0	0.0	22.4	0.5
1994	144	6	4.2	792	0	0.0	15.4	0.6
1995	122	4	3.3	846	2	0.2	12.6	0.6
1996	140	8	5.7	909	1	0.1	13.3	0.9
1997	135	6	4.4	969	4	0.4	12.2	0.9
1998	145	3	2.1	951	3	0.3	13.2	0.5
1999	170	7	4.1	850	1	0.1	16.7	0.8
2000	171	5	2.9	775	3	0.4	18.1	0.8
2001	181	14	7.7	681	6	0.9	21.0	2.3
2002	213	6	2.8	617	1	0.2	25.7	0.8
2003	168	1	0.6	576	0	0.0	22.6	0.1
Total	3831	142	3.71	21857	85	0.4	14.9	0.9

Table 4: Error rates and total expected costs under different models

Z-score and profit before tax (PBT) figures for all the firms in our sample are computed based on their last available full year accounts as at the end of September of each year from 1979 to 2003 (year t). Firms are then tracked over the next twelve months (to 30 September of year t+1) to identify those that failed. The z-score model classifies all firms with z<0 as potential failures, the PBT model classifies all firms with PBT<0 as potential failures, the proportional chance model randomly classifies firms as potentially failed/non-failed based on the average failure rate over the 25-year period and the naïve model classifies all firms as non-failures. The type I error rate represents the percentage of failed firms classified as non-failed by the respective model, and the type II error rate represents the percentage of non-failed firms classified as failed by the respective model. Overall accuracy gives the percentage of firms correctly classified in total. c_I:c_{II} is the ratio of the relative costs of type I to type II errors. Total expected costs are based on the average type I and type II error rates and ex post determined probability of failure. For illustrative purposes, we assume the cost of a type II error (c_{II}) is 1%. Assuming a constant cost ratio (c_I:c_{II}), change in the type II error cost produces a proportional change in total expected costs.

Model	**Error rate(%)**		**Overall accuracy rate (%)**	**Total expected costs (%)**			
	Type I	**Type II**		**c_I:c_{II} = 20:1**	**c_I:c_{II} = 40:1**	**c_I:c_{II} = 60:1**	**c_I:c_{II} = 80:1**
z-score	5.7	25.6	74.6	0.26	0.27	0.28	0.29
PBT	37.4	14.5	85.3	0.21	0.28	0.34	0.41
Proportional chance	99.1	0.9	98.2	0.18	0.36	0.53	0.71
Naïve	100.0	0.0	99.1	0.18	0.35	0.53	0.71

Table 5: Relative costs of misclassifications and total expected costs

Z-score and profit before tax (PBT) figures for all the firms in our sample are computed based on their last available full year accounts as at the end of September of each year from 1979 to 2003 (year t). Firms are then tracked over the next twelve months (to 30 September of year t+1) to identify those that failed. The z-score model classifies all firms with z<0 as potential failures and the PBT model classifies all firms with PBT<0 as potential failures. The type I error rate represents the percentage of failed firms classified as non-failed by the respective model, and the type II error rate represents the percentage of non-failed firms classified as failed by the respective model. Overall accuracy gives the percentage of firms correctly classified in total; c_I:c_{II} is the ratio of the relative costs of type I to type II errors. Total expected costs are based on the average type I and type II error rates and ex post determined probability of failure. For illustrative purposes, we assume the cost of a type II error (c_{II}) is 1%. Assuming a constant cost ratio (c_I:c_{II}), change in the type II error cost produces a proportional change in total expected costs.

c_I:c_{II}	Cut-off	Error rate (%)		Expected cost (%)	
		Type I	Type II	z-score	PBT
20	-1.72	22.47	15.25	0.19	0.21
30	-1.32	20.26	17.36	0.23	0.24
40	-1.03	16.74	18.87	0.25	0.28
50	-0.81	12.78	20.17	0.26	0.31
60	-0.63	10.13	21.37	0.27	0.34
70	-0.47	8.37	22.34	0.27	0.38
80	-0.34	7.93	23.84	0.29	0.41

Table 6: Firm failure probabilities by –ve z-score quintile

The z-scores of all the firms in our sample are computed based on their last available full year accounts as at the end of September of each year from 1979 to 2003 (year t). The firms are then ranked on their z-scores and for the negative z-score stocks, five portfolios of equal number of stocks are formed each year. Firms are then tracked for the next twelve months (to 30 September of year t+1) to identify those that failed. The z-score model classifies all firms with z<0 as failures.The entries in the table refer exclusively to the -ve z-score firms in our sample.

	Negative z-score quintile					
	5 (worst)	**4**	**3**	**2**	**1 (best)**	**Total firms**
Failed (%)	7.3	4.3	2.1	1.8	0.8	214
Non-failed (%)	92.7	95.7	97.9	98.2	99.2	6519
Number of firms	1356	1347	1338	1342	1350	6733
% of total failures (n = 227)	42.3	25.6	11.5	10.6	4.4	94.3