The performance of insolvency prediction and credit risk models in the UK: a comparative study

Richard H. G. Jackson[a,*], Anthony Wood[b]

[a] *Aberystwyth University, School of Management and Business, Aberystwyth SY23 3DD, United Kingdom*
[b] *University of Exeter, University of Exeter Business School, Exeter EX4 4PU, United Kingdom*

ABSTRACT

Theoretically driven, market-based contingent claims models have recently been applied to the field of corporate insolvency prediction in an attempt to provide the art with a theoretical methodology that has been lacking in the past. Limited studies have been carried out in order directly to compare the performance of these models with that of their accounting number-based counterparts. We use receiver operating characteristic curves to assess the efficacy of thirteen selected models using, for the first time, post-IFRS UK data; and investigate the distributional properties of model efficacy. We find that the efficacy of the models is generally less than that reported in the prior literature; but that the contingent claims models outperform models which use accounting numbers. We also obtain the counter-intuitive finding that predictions based on a single variable can be as efficient as those which are based on models which are far more complicated – in terms of variable variety and mathematical construction. Finally, we develop and test a naïve version of the down-and-out-call barrier option model for insolvency prediction and find that, despite its simple formulation, it performs favourably compared alongside other contingent claims models.

Keywords:
Insolvency prediction
Bankruptcy prediction
Contingent claims
Barrier options
Z-score

* Corresponding author. Tel.: +44 (0) 1970 622740.
E-mail address: richard.jackson@aber.ac.uk

1. Introduction

The prediction of corporate insolvency and the assessment of credit risk have been the subject of much academic and professional research over the last half century. As when a pebble is thrown into a lake and the shockwaves reach far beyond the point of initial impact, when a company becomes financially distressed/insolvent there are adverse consequences for its diverse stakeholder groups, such as investors, managers, employees, customers and suppliers, which impact onward into other firms, the wider economy and society. According to the president of R3 (the UK association of business recovery professionals), Steven Law (2010), 'Any future increases in corporate insolvencies are likely to affect others as Insolvency Practitioners estimate that around 27% of corporate insolvencies are triggered by another company's insolvency - the "domino effect".'

In a similar vein, a member of the Turnaround Management Association UK asserted in Cooper (2010) that:

> 'As business [sic] struggle to survive into 2010, they are likely to put increasing pressure on their suppliers. Payments will be withheld for as long as possible. If and when a company fails, it is likely that the other businesses it owes money to will get little or nothing in return. Unfortunately, the knock on effect will be that other firms will also be starved of cash and more will find themselves under financial pressure.'

The area of distress/insolvency prediction is of high economic significance in terms of the number of firms and individuals affected and the implications for investment and lending decision making. Up-to-date research is mandated by the present era of global financial difficulty which follows recent periods of financial sector institutional failure and credit crunch. Fig. 1 shows the number of UK insolvencies annually since 1960. The glut of insolvencies during the economic downturn of the early 1990's is clearly visible as is a sharp rise in the incidence of insolvencies since 2007.

***** insert Fig. 1 about here *****

As regards the prognosis for the incidence of insolvencies in the current period, R3 (2010) had the following sober message:

> 'We would still expect a spike in the number of insolvencies in the five or six quarters following a recession because many businesses that suffered during the recession find it hard to borrow as lending requirements tighten. Many of them will see "green shoots" but will not be able to fund expansion, especially if interest rates increase. And when a recession ends and assets rise in value, creditors are encouraged to move ahead with more aggressive debt collection.'

This paper evaluates a number of different methods which have been popularly employed in the prior literature to assess firm health together with some more recent approaches. In the current economic climate of global financial turmoil, we seek to assess, using data from recent insolvency cases and post-IFRS implementation, how these various methods and approaches perform for the UK. Whereas many prediction models have shown good *ex post* efficacy in past studies, our findings suggest that many of these models have more modest ability to predict insolvency or financial distress in the current economic downturn. Therefore, extreme care should be taken in relying, wholly or in part, on information obtained from a firm health-determining model for the purposes of financial or economic decision making. UK lending institutions have for some time, like their counterparts around the globe, been urged to re-expand their lending and alleviate the current credit crunch;[1] and under Project Merlin, as announced on 9th February 2011 by George Osbourne, UK Chancellor of the Exchequer, the four largest UK lending banks agreed, *inter alia*, to lend £190 billion during 2011 – with the Bank of England to monitor the achievement of lending targets. Our study provides a stark warning to lending institutions seeking to re-expand their lending portfolios as to the technologies they might employ to assess insolvency risk, a key aspect of lending risk. We concur with Abrahams and Zhang (2009), who say:

> 'A comprehensive new credit risk framework is needed—a hybrid approach that combines the best that technology can offer with expert human judgment. Such an approach can help deal with the current crisis and may lessen the extent of, or even prevent, the next one. The magnitude of the current crisis makes it abundantly clear that there is significant room – and need – for improvement in current credit assessment approaches.'

---

[1] UK Chancellor of the Exchequer, 20th May 2010: 'There is an urgent priority that is getting lending going to small- and medium-sized businesses. That is an absolute urgent priority.' [http://uk.reuters.com/article/marketsNewsUS/idUKGOVT20MAY20100520].

As part of the assessment undertaken in this paper, we consider whether or not the extensive academic endeavour in this area, with many papers and studies considering, in aggregate, a huge volume of data with diverse techniques and models, has resulted in the academic or practitioner being better able to predict corporate failure now than (s)he would have been, say, 46 years ago after Beaver (1966) published his seminal work on financial ratios and suggested an approach to distinguish between failed and non-failed firms using a single ratio. Beaver himself commented that ‘the best single ratio appears to predict about as well as the [early] multi-ratio models’ (Beaver, 1966, p. 100); and we are interested, *inter alia*, to see if this is still the case today.

This paper adds to the literature upon the relative efficiency of different models for insolvency prediction in several respects by: (i) using post-IFRS data from the UK; (ii) providing comparisons between broader range of contingent claims and accounting number-based models than are found in earlier extant studies, and using receiver operating characteristic curves as the basis for those comparisons; (iii) proposing and testing a new contingent claims insolvency prediction model; and (iv) investigating the distributional properties of the efficacy of insolvency prediction models.

The paper continues as follows. The next section reviews recent insolvency prediction literature, and the relative popularity of different insolvency prediction technologies used in academic research; the third section describes selection and development of a sample of firms and firm-level data against which to test the different models; the fourth section provides an overview of the methods and models tested within our study; the fifth section describes the means by which we compare the models; and the sixth section presents the results of the comparison. The final section concludes and provides a discussion.

2. Review of the prior literature

2.1. Evolution and recent developments

The work of Beaver (1966) seeded the modern literature on insolvency prediction with a univariate approach, treated further in Beaver (1968). Methods in the 1970s centred on a multivariate framework, with the widespread use of multivariate discriminant analysis (MDA) models and the production of the ‘Z-score’ and similar technologies (Altman, 1968; Deakin, 1972; Edminster, 1972; Blum, 1974; Diamond, 1976; Taffler & Tisshaw, 1977; Taffler, 1983). Criticisms relating to violations of the statistical assumptions underlying the MDA approach,[2] however, led researchers of the 1980s to concentrate their efforts on the development of conditional probability models, the most popular being the logit (Ohlson, 1980; Hamer, 1983; Zmijewski, 1984; Zavgren, 1985; Keasey & Watson, 1986).

Technological developments and an increase in the availability and power of computer processing allowed the insolvency prediction researchers in the 1990s to adopt a wider range of methods. Much of the work in the 1990s concentrated on artificially intelligent systems such as neural networks, genetic algorithms, case-based reasoning and recursive partitioning (Odom & Sharda, 1990; Coats & Fant, 1993; Boritz, Kennedy, & Albuquerque, 1995; Charitou, Neophytou, & Charalambous, 2004).

Despite the voluminous research, however, insolvency prediction suffered much criticism because of a lack of theoretical underpinning – giving rise to problems associated with the classical paradigm (choice of definition of firm failure, non-stationarity and instability of data, and sample selection) and the arbitrary selection of variables and modelling method.[3]

More recent modelling developments are the contingent claims models which are based on option pricing theory as set out in Black and Scholes (1973) and Merton (1974). The most popular forms of the model are derived from the European call option and the down-and-out call barrier option. Prior to the acquisition of KMV by Moody's Corp in April 2002 very little was known publicly about KMV's proprietary credit risk appraisal methodologies – which include versions of the contingent claims model. After the acquisition, a selection of papers written by KMV practitioners became publicly available to download from the Moody's KMV website,[4] and contingent claims models have since attracted a deal of interest from academics.

In its simplest form, the model assumes that the shareholders of a firm hold a European call option (EC) on the firm, the exercise price being the amount required to discharge its debt liabilities. The time to maturity of the debt is, therefore, taken to be the option period and the expiry date is taken as the point at which insolvency might occur. At expiry, the shareholders

[2] See, for example, Joy and Tollefson (1975, 1978), Eisenbeis (1977), Moyer (1977), and Altman and Eisenbeis (1978).
[3] See Balcaen and Ooghe (2006) for an excellent discussion of the problems.
[4] http://www.moodyskmv.com/research/New%20Research_sectionl.html

will either mandate discharge of the debt repayment obligation (if firm assets exceed liabilities) or mandate default on the debt and insolvency process (if firm assets are less than liabilities). The method assigns a default (failure) probability independently to each firm. Examples of the EC approach can be found in Vassalou and Xing (2004), Bharath and Shumway (2008), and Hillegeist, Keating, Cram, and Lundstedt (2004).

Several studies, including Brockman and Turtle (2002) and Reisz and Purlich (2007), derive default probabilities by extending the EC model into a barrier options framework. The model values the option as a down-and-out call (DOC). Debt holders are deemed to own a portfolio of risk-free debt and a DOC option on the firms' assets which can be exercised should the value of the firm fall below a predetermined legally binding barrier.

Comparisons of the performance of contingent claims models with that of traditional accounting-number based models forms the backbone of much of the empirical analysis found in the recent research and such comparisons have yielded a variety of results. Tudela and Young (2003) find that their contingent claims model out-performs models which use accounting data alone and that, when the output of their model is combined with accounting data, accuracy is only slightly improved. Hillegeist et al. (2004) compare their EC model with two accounting number-based methods (being the 'Z-score model of Altman, 1968 and the logit model of Ohlson, 1980) and find that the contingent claims model carries more information than the accounting number-based comparators. They argue that, since accounting statements are prepared on a going-concern basis, they are, by design, of limited use in predicting bankruptcy. Bharath and Shumway (2008) construct a naïve version of the EC model in order to avoid some of the estimation complexity found in Hillegeist et al. (2004), and find little deterioration in model performance. Reisz and Purlich (2007) compare both EC and DOC models with the Altman Z-score model and find, in contrast to the findings of Hillegeist et al. (2004), that the Z-score model out-performs both contingent claims approaches in terms of receiver operating characteristic curves for a one-year ahead forecast. For the UK, Agarwal and Taffler (2008) find similarly to Reisz and Purlich (2007) when comparing the Taffler (1983) accounting number-based model (the Taffler Z-score) with EC models of both Hillegeist et al. (2004) and Bharath and Shumway (2008). Agarwal and Taffler (2008, p. 1,542) suggest that '...traditional accounting ratio-based bankruptcy risk models are, in fact, not inferior to KMV-type option-based models for credit risk assessment purposes ... The apparent superiority of the market-based model approach claimed by

Hillegeist et al. (2004) reflects the poor performance of their comparator models, not a particularly strong performance by their option-pricing model.'

2.2. A wide diversity of approaches

The academic literature contains a vast array of techniques which have been posited over the last five decades to 'predict' corporate insolvency. Our review of the literature has identified 25 different methods (many more if variations within method are counted) which are shown as the abscissa labels in Fig. 2. These range from the univariate and multivariate discriminant analysis of the 1960s and 1970s, through the logit and probits of the eighties and the artificial learning models of the 1990s, to the recent crop of contingent claims models in the 2000s.

A number of papers compare and contrast these various methods and techniques, for example Scott (1981), Zavgren (1983), Laitinen and Kankaanpää (1999) and Balcaen and Ooghe (2006). A meta-analysis by Aziz and Dar (2006) considers a number of methods and techniques from a sample of 89 empirical studies, comparing the frequency of usage (by academics) and efficacy of sixteen different insolvency prediction methodologies, and assigning each of them to one of three categories: 'statistical models'; 'artificially intelligent expert systems (AIES)'; and 'theoretical models'. The paper shows MDA and logit to have been the most popular models for study amongst academics.

Fig. 2 shows the frequency of academic investigation of each of the 25 corporate insolvency prediction technologies found in our literature review, grouped according to the Aziz and Dar (2006) categories. Having surveyed over 350 papers, a larger sample than was considered by Aziz and Dar, the relative frequency of academic investigation of techniques that we find is in accordance with their results.

***** insert Fig. 2 about here *****

The selection choice of technologies for comparative study in this paper is based on three factors: the frequency of study and reported efficacy in the prior academic literature; relevance to the UK; and the desire to incorporate and compare contingent claims models and their traditional accounting number-based counterparts. We arrive at a sample of thirteen models representing five technologies. The details of the models and reasons for inclusion are described in a later section.

3. Sample of firms and collection of firm-level data

This section describes the development of samples of failed and non-failed firms for our study, and the collection of data upon those firms.

3.1. Population and sample selection

As regards failed firms, this study takes the population of all London Stock Exchange (LSE) listed non-financial firms recorded on the London Share Price Database as failing between 30th September 2000 to 31st December 2009, reduced only by the exclusion of firms for which insufficient data is available to allow application of the insolvency prediction models under consideration. Failed firms were identified from the LSPD general descriptive record G10 'Type of Death' by codes 7, 16 and 20.[5] Firms with financial industry codes[6] are excluded from the population because of their differing capital structure, earnings profile and reporting. The final number of failed firms considered in the study is 101.

The study uses accounting and market data upon the failed companies and a sample of LSE-listed non-financial non-failed firms, the sample period for such data being 1st January 2000 to 30th September 2009. All firm specific information is obtained through LSPD and Thomson One Banker.

The sample of non-failed firms was selected from a non-financially coded population which was deemed to be 'alive' at 30th September 2009 with LSPD code G10 equal to zero. For this study a non-failed sample containing 2,244 eligible firms was identified for potential analysis. The resulting non-failed group provides 6,494 firm-years of data within the sample period where sufficient data is available to allow application of the insolvency prediction models under consideration.

Insolvency prediction literature has traditionally relied on matched samples of failed and non-failed firms, as in, for example, Beaver (1966), Altman (1968), Libby (1975), Taffler (1983), Keasey and McGuinness (1990), and Charitou et al. (2004). The paired sampling approach is a legacy of the discriminant analysis methodology frequently used in the 1960s and 1970s, where failed firms were matched on size and industry with non-failed mates.

[5] Representing, respectively: *creditor liquidation*, *receiver appointed*, and *in administration*.
[6] LSPD (G17) codes 8XX and 8XXX.

Paired sampling is not, however, a necessity for this work. Stein (2002) posits that the variable accuracy of any prediction model is ‘primarily driven by the number of defaults, rather than the total number of observations’. This is supported by Falkenstein, Boral, and Carty (2000), who suggest that the hazard model of Shumway (1999) outperformed others in their evaluation because of the large number of defaults – 300 – in the sample.

Current insolvency prediction and credit risk papers utilize the abundance of financial databases and related search tools to include a far greater number of non-failed firms in relation to the failed ones, more realistically representing the percentage of failures in the actual population. These technological advances mean that all listed firms have the potential to be analysed in greater detail. Moody’s ‘RiskCalc for Public Companies’, for example, uses a database of 1,406 failed and 13,041 non-failed firms for their proprietary adaptation of the contingent claims model in the US. In the UK, the number of insolvencies is far lower than that in the US, the overall population of UK firms being itself much smaller. Agarwal and Taffler (2008) use data from 2,006 firms including 103 failures over a fifteen year period in the UK; this is far fewer than Hillegeist et al. (2004), who studied 756 failures and 14,303 non-failed firms over an almost identical period in the US. Table 1 compares sample sizes from a number of the studies cited in this paper.

***** insert Table 1 about here *****

3.2. Data collection

Firm name, industry classification, death date, death type and other listing details are collected from LSPD. Accounting data is sourced from Thomson One Banker and market data collected from both Thomson One Banker and LSPD. The risk-free rate of return is calculated annually using the nominal yield on a 12-month UK treasury bill as at 30th September each year.

3.3. Structure of annual samples

Our empirical analysis involves the consideration on the same date each year of a set of all firms from our sample (which fail within twelve months and those which do not) for which pertinent data is available in respect of that year. In the prior literature, such sets have been referred to as annual ‘portfolios’.

We create portfolios for years 2000 to 2009 inclusive. The choice of portfolio date (the same from year to year) is somewhat arbitrary. We adopt 30th September, following Agarwal and Taffler (2008), which aids comparability with that study and between the different models which we test. Each portfolio is constructed using market data up to and including 30th September. Accounting data, however, is taken only from annual reports made up to 30th April of the same year or before, to duplicate the information conditions which would be likely to have faced a real-world user of the models on 30th September of the year concerned.

3.4. Training and validation

For some of the prediction models which we test, both training and validation samples are required. There are various alternative forms of approach to validation (e.g., random sample cross-validation, *k*-fold cross validation, Lachenbruch jack-knife). In order to reflect the practices of construction and use of the models by real-world users, we first split the sample through time. Training data comes from portfolios 2000-2005 inclusively, and the validation data (on which all models, irrespective of whether or not they require training, are tested) is derived from portfolio years 2006-2009.[7] Table 2 presents the number of firm-year observations employed in these training and validation samples.

***** insert Table 2 about here *****

Additionally, we generate 10,000 random splits between training and validation samples. In each case the number of both failed firm-year observations and non-failed firm year observations in the training and validation samples are as in Table 2 but with the allocations now being made on a random basis, rather than according to chronology. This provides us with the opportunity to investigate variation in efficacy of the prediction models resulting from variation in allocation of observations between the training and validation samples.

[7] Obtaining an acceptable sample size (in respect of number of failed firms) to allow use of post-IFRS implementation data for both training *and* validation samples is not yet possible.

4. Overview of selected models

This section presents an overview of the methods and models tested within this study. For a more detailed discussion of the various methods, we suggest Zavgren (1983), Crosbie and Bohn (2003), Balcaen and Ooghe (2006), and Reisz and Purlich (2007).

4.1. Single variables

We include three single variable 'models' in our analysis: cash-flow to total debt (which we designate model BV); firm size (model SIZE); and book-to-market value ratio (model B/M). Cash-flow to total debt was determined by Beaver (1966) to be the best performing ratio, amongst those he tested, for the prediction of failure. Re-examination of this ratio is infrequent in recent literature as most studies focus on an array of multivariate and non-linear approaches.[8] We include Beaver and his best performing ratio to investigate how one of the simplest and earliest posited methods of insolvency prediction compares with the latest efforts of academics and practitioners working with modern computer processing power. That is, to see if developments over the last 46 years have significantly advanced our ability to separate the failing from the non-failing firms.

Vassalou and Xing (2004) find empirical evidence that 'both size and book-to-market exhibit a strong link with default risk'; and Agarwal and Taffler (2008) use these variables as benchmarks against which to compare multivariate models. Size is intuitively appealing as a factor in a firm's chances of surviving financial setbacks, since high equity reserves or asset bases act as a buffer to enable continued debt servicing. For example, BP plc, one of the largest companies listed on the FTSE, saw its share-price collapse in 2010 following ecologic and political blunders surrounding the Deepwater Horizon disaster. The size of BP has ultimately ensured its survival, despite a $20 billion spill response fund it was ordered to create. In contrast, one the highest profile UK companies to enter insolvency process in the last five years was Woolworths plc. Although one of the UK's largest and longest established high street retailers, with an asset base of just 1% of BP's, Woolworths could not weather the onset of recession and other problems. We measure size using the natural logarithm of GDP-deflated market value of equity as a proxy.

[8] Although Zavgren (1983), for example, includes discussion of Beaver's ratio.

Our inclusion of the book to market ratio allows, in essence, investigation of the extent to which market capitalisation of future earnings alone is useful in predicting failure.

4.2. Multivariate discriminant analysis: z-score models

Z-score models consist of a linear combination of variables estimated using MDA, which classifies observations into one of a number of pre-specified categories or groups.[9] In the present context there are two groups, being failed firms and non-failed firms. The resulting discriminant function is of the form:

$$Z_i = \alpha + \beta_{j1}X_{i1} + \beta_{j2}X_{i2} \ldots\ldots + \beta_{jn}X_{in} \quad (1)$$

where Z is the 'score' for firm i; α is a constant; the X_{ij} are the attributes (ratios, categorical or qualitative variables) for firm i; and the β_j are coefficient estimates for each attribute. Scores allow for an ordinal ranking of firms, the higher the score usually denoting the better the predicted solvency.[10] The adoption and application of a cut-off point in score may then be used to divide firms between those predicted to fail and those predicted to survive but the choice of cut-off point is somewhat arbitrary.

Brief discussion of three methodological issues in the use of MDA models in the present context is necessary. First, many researchers have chosen cut-off points in order to attempt to minimise the *overall* error rate of the model - which may result in the misleading or incorrect estimates of model accuracy. Differing economic impacts associated with type I errors (incorrect classification a failing firm as a non-failing firm) and type II errors (incorrect classification of a non-failing firm as failing) have often been acknowledged (Edmister, 1972; Eisenbeis, 1977; Deakin, 1977; Zavgren, 1983). Some studies, such as Altman, Haldeman, and Narayanan (1977), investigate differing error costs, and others, such as Agarwal and Taffler (2008), look at economic value when misclassification costs are different. Second, the classic insolvency prediction paradigm is that the dependent variable is dichotomous, treating

[9] A full derivation of the discriminant process, function, and maximum likelihood estimation technique can be found in Dhrymes (1974, pp. 65-77). The functional form does not have to be linear (e.g., Altman et al., 1977 use quadratic discriminant analysis). Linear discriminant analysis is, however, by far the most popular method in the prior literature and, therefore, we use 'MDA' to refer to the linear method.

[10] Studies such as Ooghe, Joos, De Vos, and Bourdeaudhujj (1994) define their model alternatively, higher scores denoting higher probability of failure. Later in this paper, we change the signs on our tested Z-score models not only to match such ranking systems but also directly to compare these to other methods which, instead of rankable scores, provide probability of failure estimates. A later section of the paper explains this transformation in more detail.

failure as being discrete, non-overlapping and identifiable and thus does not reflect the true nature of financial distress and the various insolvency procedures which may or may not ensue. Third, two statistical requirements of the MDA model are of multivariate normality and that the category groups into which observations are separated have identical variance-covariance matrices. These statistical requirements are rarely satisfied by the data, and so concerns as to bias pertain. A consequence of the foregoing is that the MDA modelling technique is often applied in an inappropriate way, with the resulting models being unsuitable for generalization (Joy & Tollefson, 1975; Eisenbeis, 1977; Balcaen & Ooghe, 2006).

We examine two forms of the Z-score model, the Altman (1968) Z-score and the Taffler (1983) Z-score,[11] both re-estimated using our own data set, designating the models AZU (Altman, Z-score, updated) and TZU (Taffler, Z-score, updated) respectively. Re-estimation is vital, since the original Altman model was derived from US data over the period 1946-1965; and the Taffler Z-score was based upon 1968-1976 data, albeit from the UK. The models are included in our study owing to their popularity in prior literature, past reported efficacy and, for the latter, UK relevance. The training samples for the re-estimations (and the testing/validation samples) are as described in the previous section.

4.3. The logit model

The logit model is a conditional probability model which uses the non-linear maximum log-likelihood technique to estimate the probability of firm failure under the assumption of a logistic distribution. The parameter estimates are obtained using the logit model's maximum likelihood method as derived in Gujarati (2003). The resulting model is of the following form:

$$P_i = E(Y = 1|X_i) = \frac{1}{1+e^{-(\alpha+\beta_{j1}X_{i1}+\beta_{j2}X_{i2}\ldots\ldots+\beta_{jn}X_{in})}} = \frac{1}{1+e^{-Z}} \qquad (2)$$

where P_i is the probability that firm *i* will fail given a vector of attribute variables X_{ij} (ratios, categorical or qualitative variables) for firm *i*, and the β_j are parameter estimates.

Following the majority of prior literature, we refer to the model as being constructed using a dummy variable taking the value of 1 for failed and 0 for non-failed. Given the underlying logistic function of the model, an extremely healthy (weak) company, as compared to a firm

[11] The discriminant functions, including variable definitions, for these two models are set out in the Appendix to this paper.

of average financial health, must experience a proportionally larger deterioration (amelioration) of its attributes in order to deteriorate (ameliorate) its logit score (Laitinen & Kankaanpää, 1999).

Adopting the logit model circumvents some of the statistical assumptions violated by MDA (Edmister, 1972; Eisenbeis, 1977; Altman & Eisenbeis, 1978; Joy & Tollefson, 1975; Joy & Tollefson, 1978; Ooghe et al., 1994). In particular, the logit model neither requires multivariate normally distributed variables, nor does it rely on equal variance-covariance matrices of the two classification groups (Ohlson, 1980; Zavgren, 1983). The logit model does, however, like MDA, rely on two basic assumptions inherent in the classical prediction paradigm: the dependent variable is still dichotomous, treating failure as being discrete, non-overlapping and identifiable, thus not reflecting the true nature of financial distress and the various insolvency procedures; and the input variables are still chosen arbitrarily.

We test one particular version of the logit model, that of Ohlson (1980),[12] re-estimated using the training samples from our own data set, and designate the model OLU (Ohlson, logit, updated). As with the Z-score models, re-estimation is vital – since the original Ohlson model was derived from US data from the period 1970-1976. Inclusion of this model is based on its popularity and influence in the prior literature (e.g., Zavgren, 1983; Begley, Ming, & Walls 1996; Hillegeist et al., 2004).

4.4. Artificial neural networks

Neural networks (NNs) were originally conceived in the 1940s and 50s (McCulloch & Pitts, 1943; Hebb, 1949; Rosenblat, 1957) as a way of mimicking the function of the human thought process. They have been applied extensively to an increasingly wide variety of business areas, including financial forecasting, credit analysis, bond ratings, bankruptcy prediction and fraud detection (Charitou et al., 2004).

A NN consists of a number computational functions, called 'neural nodes' or 'neurons', organised in a particular structure with particular inter-connections which are dependent on the designated application. We employ a typical NN structure, a feed-forward network structure, in a basic form of which has three layers of neurons: the input layer; hidden layer; and output layer, as shown in Fig. 3. The connections between these layers flow in one

[12] The specification for this model, including variable definitions, is set out in the Appendix to this paper.

direction from the input node through the network in a forward ‘direction of activation propagation’. Data provided at the input nodes is sent via weighted parallel connections to provide input data to the nodes in the hidden layer which is then transformed using a non-linear, sigmoid function and sent to the output layer. For the transformation, we employ the logistic function as described above. Thereafter, a similar process at the output layer yields a single output variable. Mathematically, the operation executed by a hidden or output neuron with *n* inputs is described by the following:

$$output = f\left(\sum_{i=1}^{n} w_i \times input_i\right) \tag{3}$$

where w_i represents the weight given to input *i*, and *f* represents the logistic transformation.

***** insert Fig. 3 about here *****

The NN is ‘trained’ into a useful network by ‘supervised learning’, using a training algorithm and data from a training sample. We use the backward propagation of errors (or ‘back propagation’) technique for our training algorithm. In essence, the weights on inputs to each node are adjusted using an iterative process in order to ever more closely map inputs to known outputs in the training sample. In the present context, firm characteristics (ratios, categorical or qualitative variables) are the inputs, and firm failure is the output. A full derivation of the back propagation algorithm can be found in Rojas (1996).

Unlike the MDA and logit approaches, the NN has the advantage that it does not rely on any prespecification of a functional form, nor, in contrast to the MDA and logit approaches discussed earlier, is it restricted in its assumptions regarding the characteristics and statistical distributions of variables. For these reasons, and the popularity of NNs in the prior literature, we include this technology in our analysis. NNs are not, of course, a methodological panacea. Two commonly cited issues with their use are first, the tendency for ‘overfit’ of the network to the training sample, with consequential issues for model usage out of sample and, second, the difficulty in understanding or interpreting the network model.

We apply the NN methodology described above to form three different models based upon, in turn, the variables employed by Altman (1968), Ohlson (1980) and Taffler (1983), designating these models AZN, OLN, and AZN respectively. By comparing each model estimated and applied as originally formulated with NN models employing the same input parameters, we seek to evaluate further the ongoing usefulness of those parameters as predictors of corporate insolvency while abstracting from the caveats and restrictions of the earlier functional forms.

4.5. The European call contingent claims model

The EC model is used to measure the default probability of a firm. It is based on the work of Black and Scholes (1973) and Merton (1974), employed and modified by Moody's Rating Methodology as described in Falkenstein et al. (2000), Sobehart and Stein (2000) and Crosbie and Bohn (2003).

The model views equity as a European call option on a firm's assets with a strike price equal to the face value of its debt liabilities.[13] The option expires at time *T* when the debt matures, at which point the equity holders either: (i) exercise their option and pay off the debt if the value of the firm's assets is greater than that of its liabilities; or (ii) let the option expire, if the assets are not sufficient to cover the cost of the maturing debt. If the option is left to expire, insolvency is assumed to ensue and the residual claim to equity is assumed zero. The contingent claims model determines the probability of the two outcomes. McDonald (2002, p. 604), shows that the probability of failure may be calculated as follows:

$$Prob = N\left[-\frac{ln\left[\frac{V}{X}\right] + \left(\mu - \delta - \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}}\right] \qquad (4)$$

where *X*, the face value of long term liabilities, and δ, the annual dividend rate, are directly observable; *T*, time to expiry, is taken to be one year in the present context; and *V*, the market value of the firm's assets, σ, the asset volatility, and μ, the firm's expected return, are not directly observable and must, therefore, be estimated.

We proceed to summarise two versions of the EC model which are evaluated within our study. These versions differ in their approach to the estimation of the unobservable parameters *V*, σ and μ, and in treatment of the *X* value.

[13] All liabilities are assumed to be zero-coupon bonds following Black and Scholes (1973) and Merton (1974). Hillegeist et al. (2004) point out that, further to Merton (1974), 'Subsequent studies have incorporated more realistic assumptions, such as allowing for debt covenants (e.g., Black and Cox, 1976) and multiple classes of debt (e.g., Geske, 1977). We leave an evaluation of alternative option pricing models to future research'.

4.5.1. Hillegeist et al. (2004) approach

The value of equity, V_E , as a call option on the value of the firm's assets is:[14]

$$V_E = Ve^{-\delta T}N(d_1) - Xe^{-rT}N\left(d_1 - \sigma\sqrt{T}\right) + \left(1 - e^{-\delta T}\right)V \tag{5}$$

where N(.) represents the cumulative standard normal function, *r* represents the risk free rate of interest as proxied by 12 month UK Treasury Bond yields, $d_1 = \frac{ln\left[\frac{V}{X}\right] + \left(r - \delta + \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}}$ and all other variables are as previously defined.

Here, the optimal hedge equation is:

$$\sigma_E = \frac{\left[Ve^{-\delta T}N(d_1)\sigma\right]}{V_E} \tag{6}$$

The values V and σ may be found by solving equations (5) and (6) simultaneously. We use a modified version of the SAS code as provided by Hillegeist et al. (2004, pp 30-31) with initial values of V and σ computed according to Bharath and Shumway (2008). The values of X are proxied by total liabilities, departing from the Moody's method that uses current liabilities plus half of long-term debt. Hillegeist et al. suggest that lowering the value of liabilities mechanically reduces the probability of failure estimates and slightly weakens the model.

The expected return on the firm's assets, μ, is proxied by the actual return on assets for the previous year, bounded below by the risk-free rate and bounded above by 100%. Taking this with values derived from equations (5) and (6) into equation (4) provides the Hillegeist et al. probability of default.[15] We designate this model HKCL.

[14] The call equation is modified to include dividends, reflecting that the dividend stream flows to equity holders.
[15] Moody's methodology differs here in that the defaults are not taken from an assumed normal distribution, but instead mapped to historical default frequencies from their large proprietary database.

4.5.2. The Bharath and Shumway (2008) 'naïve' approach

The naïve EC model presented by Bharath and Shumway (2008), which we designate BS, simplified the foregoing by estimating V and σ using the following equations:

$$V = V_E + X \tag{7}$$

$$\sigma = \frac{V_E}{V}\sigma_E + \frac{X}{V}\sigma_D \tag{8}$$

$$\sigma_D = 0.05 + 0.25\sigma_E \tag{9}$$

where σ_D represents the volatility of the firms debt, naïvely deduced by a linear transformation of the volatility of equity, σ_E - taken as standard deviation of daily log returns over the previous twelve months; and other variables are as previously defined, albeit X and μ are now estimated as follows.

Bharath and Shumway approximate the value of liabilities, X, to be current liabilities plus one half of long term debt, following Moodys' methodology. Reisz and Purlich (2007) present that within a barrier option framework, the barrier that triggers insolvency is more likely to be lower than the face value of long-term liabilities, therefore, current liabilities plus one half of long term debt may be a more realistic approach for a call option model. The expected return on the firm's assets, μ, is estimated simply by using the prior year stock return bounded between the risk-free rate and 100%. According to Agarwal and Taffler (2008), a problem with this bounded variable is that 'even if realized return on equity is a good proxy for expected return on equity, it will be a good proxy for expected return on assets only if the expected return on debt is the same for all the firms.' More problematic may be that, in a real world setting, the prior year stock return may be negative, particularly within a sample period enveloping a period of economic downturn, thus setting a floor at the risk-free rate does not reflect the possibility of equity degradation. Agarwal and Taffler (2008) show, however, that the expected return generating model does not significantly affect the outcome of the probability estimates.

4.6. The down-and-out call barrier option contingent claims model

The barrier option framework extends the EC model and treats the underlying value of firm equity to be treated as a DOC option. Debt holders own a portfolio of risk-free debt and a DOC option on the firms' assets which can be exercised should the value of the firm falls below a predetermined barrier. The implied firm-specific barrier level is estimated by using market values of firm traded equities (or an appropriate proxy) which can be achieved by adopting equation (10) and solving for the barrier, *B*. Equation (10) shows the standard DOC formula under the condition that the barrier is set at, or below, the strike-price/liability value, *X*.[16]

$$V_E = DOC = Ve^{-\delta T}N(d_1) - Xe^{-rT}N\left(d_1 - \sigma\sqrt{T}\right) - \left[Ve^{-\delta T}\left(\frac{B}{V}\right)^{\frac{2(r-\delta)}{\sigma^2}+1} N(d_1^B) - Xe^{-rT}\left(\frac{B}{V}\right)^{\frac{2(r-\delta)}{\sigma^2}-1} N\left(d_1^B - \sigma\sqrt{T}\right)\right] \quad (10)$$

where $d_1^B = \frac{ln(B^2/VX)+\left(r-\delta+(\sigma^2/2)\right)T}{\sigma\sqrt{T}}$ and all other variables are as previously defined and *T* is again taken as one year.

The setting of the barrier (higher than, lower than, or at the liability value) is a matter of debate, of which a good discussion is provided by Reisz and Purlich (2007). For example, companies which exhibit a low credit rating, high monitoring costs or long term illiquid assets, may see debt-holders set a barrier above the liability value. Given a definition of insolvency as being $V < X$, however, we would not anticipate that a barrier set above liabilities would trigger insolvency. Rather, we would expect that the position of the barrier would be set at, or below, the value of debt to be serviced. Moody's practitioners Crosbie and Bohn (2003) find that

> '... in general firms do not default when their asset value reaches the book value of their total liabilities. While some firms certainly default at this point, many continue to trade and service their debts. The long-term nature of some of their liabilities provides these firms with some breathing space. We have found that the default point, the asset value

[16] See Brockman and Turtle (2002) and Reisz and Purlich (2007) for details of possible variations, and an excellent discussion and formulation of barrier options framework. The DOC formula is often presented as containing a term to take into consideration a possible rebate which may be attributable to the shareholders should the option expire. As with similar studies, we assume a zero rebate - so the term collapses. Note that if both the rebate and the barrier are set to zero, then the equation further collapses into the non-dividend EC option model.

at which the firm will default, generally lies somewhere between total liabilities and current, or short-term, liabilities.'

As with the first EC model described above, the values V and σ are determined by solving the call option equation (5) and the optimal hedge equation (6). Several papers follow Moody's and set liabilities, X, equal to current liabilities plus half long-term debt. [17] Others, including Hillegeist et al. (2004), stick more closely to the Black-Scholes-Merton theory by using total liabilities for X, and we do the same. Once B has been backed out from equation (10), the risk-neutral probability of default can be calculated using equation (11), for which a concise derivation may be found in Reisz and Purlich (2007, Appendix A, pp.123-129). We designate this model DOC.

$$Prob = N\left[\frac{ln\left(\frac{B}{V}\right)-\left(\mu-\delta-\frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}}\right] + \left(\frac{B}{V}\right)^{\frac{2(\mu-\delta)}{\sigma^2}-1} N\left[\frac{ln\left(\frac{B}{V}\right)+\left(\mu-\delta-\frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}}\right] \tag{11}$$

where all variables are as previously defined and T is taken as one year.

4.6.1. A new naïve DOC model

It may be argued that a large part of the interest which has surrounded the Altman Z-score, further to its claimed predictive ability, is its sheer simplicity. Five variables and a simple linear formulation build to an insolvency prediction model which is held in mind by all academics and students in the field. In the spirit of Bharath and Shumway (2008), who proposed a naïve version of the EC option model, we now turn to formulation of a simplified DOC model for the estimation of within-one-year probability of failure. This is more easily applicable than the foregoing DOC model, yet retains the essential theoretical basis. We first follow Bharath and Shumway in simply estimating V and σ using equations (7) to (9), but now taking X as the book value of total debt. We then calculate the firm-specific barrier, not by adopting equation (10), but rather setting the barrier to the same level as the firm's total liabilities – which is consistent with a balance sheet view of insolvency, i.e., $V < X$. Further, we assume no dividends, zero rebate, costless insolvency proceedings, and set the return on assets equal to the risk-free rate. The new naïve model, which we designate JW_DOC, is summarised as follows:

[17] For example: Vassalou and Xing (2004), Bharath and Shumway (2008), Reisz and Purlich (2007).

$$JW\_DOC = N\left[\frac{ln\left(\frac{X}{V}\right)-\left(r-\frac{1}{2}\sigma^2\right)}{\sigma}\right] + \left(\frac{X}{V}\right)^{\frac{2r}{\sigma^2}-1} N\left[\frac{ln\left(\frac{X}{V}\right)+\left(r-\frac{1}{2}\sigma^2\right)}{\sigma}\right] \quad (12)$$

where variables are as previously defined and from which we omit *T*, this being taken as one year.

4.7. Summary of characteristics of models to be tested

As is evident from the second section of the paper, and from the foregoing in this, the fourth section, there is enormous variety in both the nature of insolvency prediction models and in the terminology used to describe those models. In the interests of clarity, Table 3 presents a summary of the models selected for testing within this study.

***** insert Table 3 about here *****

We proceed to evaluate all the models individually and compare them according to different measures. In terms of model groups, the key comparisons which we draw out in the results and conclusions sections of this paper area: (i) accounting numbers-based models versus contingent claims models; and (ii) single variable models versus multivariate models.

5. Model evaluation and comparison

5.1. Comparability

In order that the continuous Z-score and single ratio models may directly be compared to both the logit and contingent claims probability measures, we follow both Hillegeist et al. (2004) and Agarwal and Taffler (2008) by using the logistic cumulative function $p = e^{(z\ score)}/(1 + e^{(z\ score)})$. This transformation is not a panacea, as noted by each of the aforementioned studies. Maddala (1983) shows that discriminant scores should be converted into probabilities using a linear probability model after suitable transformation through regression sum of squares. McFadden (1973) shows, however, that the MDA and logit approaches are closely related under normality assumptions and therefore the use of the logistic cumulative transformation is valid.

5.2. Receiver Operating Characteristic (ROC) curves

For our model comparison we assess each model's predictive ability using the ROC curve. Traditional tests for model accuracy have relied on the minimising of type I and type II errors based on an arbitrary cut-off point chosen by the researcher. Table 4 shows the form of a typical classification table, similar to that first used by Altman (1968). The problem with this classification method is the single cut-off point. The ROC curve addresses this issue by evaluating the model's performance over the whole range of possible cut-off points.[18]

***** insert Table 4 about here *****

Our construction of ROC curves is as found in Gönen (2006). Firm-year observations from each portfolio are ranked from the highest probability of failure to the lowest. For each of 101 possible cut-offs in probability, ranging from 1.00 to 0.00 on steps of 0.01, the number of *TP*, *FP*, *TN* and *FN* outcomes are be observed. Then for each of the 101 possible cut-offs, the true positive rate (TPR) and the true negative rate (*TNR*) are deduced. The *TPR* is calculated as *TP* / (*TP*+*FN*), and is often referred to as the 'sensitivity' or 'hit rate'. The *TNR* is calculated as *TN* / (*TN*+*FP*) and is referred to as the 'specificity'. The ROC curve is then plotted as sensitivity against 1-specificity (also known as the 'false positive rate' or 'false alarm rate') as cut off probability varies. The model's overall accuracy is then deemed to be represented by the area under the ROC curve, θ.[19] The value of θ is expressed as a percentage; and a value of 100% indicates a perfect model in that, for every possible cut-off, sensitivity and specificity are both 1, i.e., the ROC is made up of a series of superimposed plot points at the top left of the ROC diagram. A model which has no predictive ability in excess of a random assignment between failed and non-failed would have θ = 50%.

The area under the ROC curve is estimated using the trapezoid rule, which Hanley and McNeil (1982) show is 'virtually identical' to the Wilcoxon test statistic and an unbiased estimator of the true smoothed area under the curve. As our ROC curves are constructed using percentile values of specificity and our probability estimates are on a continuous scale, the trapezoid rule for calculating the area is empirically sound. The standard error of the area

[18] ROC curves have traditionally and most commonly been used in the field of medicine for assessing various diagnostic procedures and treatment techniques.
[19] Stein (2002) provides an excellent examination and interpretation of the meaning of the area under a ROC curve.

under the ROC curve, $SE(\hat{\theta})$, is calculated following Hanley and McNeil (1982) as per equation (15).

$$SE(\hat{\theta}) = \sqrt{\frac{\hat{\theta}(1-\hat{\theta})+(n_F-1)(Q_1-\hat{\theta}^2)+(n_N-1)(Q_2-\hat{\theta}^2)}{n_F n_N}} \tag{15}$$

where n_F and n_N are the number of failed and non-failed firms in the sample respectively; Q_1 is the probability that two randomly selected failed firms will both be classified as having a higher probability of failure than a randomly selected non-failed firm ($Q_1 = \hat{\theta} \div (2 - \hat{\theta})$); and Q_2 is the probability that one randomly selected failed firm will be classified as having a higher probability of failure than two randomly selected non-failed firms ($Q_2 = 2\hat{\theta}^2 \div (1 + \hat{\theta})$). The standard error is used to determine the *t*-statistic for any significant difference in the model's accuracy over a model with no predictive ability.[20]

The accuracy ratio, *AR*, of each model, following Englemann, Hayden, and Tasche (2003), is a linear transformation of the area under the ROC curve: *AR* = 2(θ - 0.5). The maximum possible value for *AR* is 1, which would indicate a perfect model.

6. Results

We present first, in Section 6.1, results based on our chronological split between training and validation/testing samples (with training data from portfolios 2000-2005 inclusive, and validation data from portfolio years 2006-2009). This split was performed in order to reflect the practices of construction and use of the models by real-world users. We then proceed in section 6.2 to present findings as to the distributional properties of θ for each model, derived by a 10,000-iteration Monte Carlo approach to training and validation sample splits.

6.1 Results based on chronological split between training and validation samples

The mean predicted probability of firm failure as between failed on non-failed firms from each model is presented in Table 5. All the models give a higher mean predicted probability

[20] A model with no predictive accuracy, a random choice model, is represented by a 45degree line on the ROC diagram and therefore has an area of 0.5.

of failure for the group of actually failed firms than they do for the non-failed firms[21]; and this is significant at the 1% level for nine out of thirteen models - the exceptions being OLN, where the significance is only at the 5% level; and AZN, TZN and OLU, where there is no significance at generally acceptable levels.

***** insert Table 5 about here *****

The measure of how closely a model's predicted probability of failure conforms to the actual failure rate is known as calibration. The actual failure rate within our sample is 1.7%, being 62 of 3,592 firm-years in the validation sample. This percentage is much lower than that predicted by many of the accounting number-based models tested here (excluding neural network versions), indicating that these models are not particularly well calibrated.

Far better calibrated in general are the contingent claims models, with models HKCL and DOC predicting failure in 1.1% and 1.2%, respectively, across all firms; and their naïve counterparts, models BS and JW_DOC, predicting 2.4% and 4.7% respectively.[22] The failure probabilities provided by our European call models HKCL and BS are close to (yet marginally higher than) the 0.96% and 2.12%, respectively, reported by Agarwal and Taffler (2008). Therefore, the financial turbulence and economic recession during our validation sample period does not seem to have had any dramatic impact on the market data-based contingent claims models' failure probability calibration. The single variable SIZE model (itself based on market price data) is reasonably, but no so, well calibrated, predicting failure of 6.6% of firms.

Stein (2002, p.7) suggests that the calibration of a model can be achieved or improved by mapping its score to an empirical probability of default using historical data and adjusting for the difference. Without Moody's proprietary database, however, this is somewhat problematic. Re-calibration of the models is not a necessity here, however, because it does not impact on the overall accuracy of the individual model to discriminate between failed and non-failed firms and, hence, overall model power.[23]

[21] Albeit this is only apparent in the fifth decimal place for model AZN.
[22] Also, model OLU predicts a 1.5% failure rate across the sample.
[23] For a more detailed discussion of power, accuracy, and calibration refer to Stein (2002).

6.1.1. Receiver Operating Characteristic (ROC) curves

The ROC results are presented in numerically in Table 6 and graphically in Fig. 4. The accuracy ratio of the single variable B/M model is a low outlier at 0.15; with the accuracy ratios of the rest of the models being in the range 0.31 to 0.68. Again, we see the contingent claims models performing well in comparison with the accounting number-based models, with models HKCL and DOC at 0.65 and 0.64 respectively; and their naïve versions at 0.68 and 0.66 respectively. The single variable SIZE model, with an accuracy ratio of 0.61, stands above the accounting number-based models, with the exception of OLN whose accuracy ratio is 0.62.

***** insert Table 6 about here *****

and

***** insert Fig. 4 about here *****

Focusing on the area under the ROC curve, the single variable B/M model is the weakest, having θ = 57.7%, which is significantly different from 50% at only the 5% level. This is some distance from 68% θ reported for B/M by Agarwal and Taffler (2008), based on earlier UK data.

All the other models which we test, however, have θ significant at the 1% level; indicating, broadly, success in our setting of the variables and technologies employed in the prediction of firm failure. The multivariate accounting number-based models, AZU, TZU and OLU (implemented using original variables and technologies, re-estimated on our own training data set) have θ of 68.9%, 65.3% and 78.1% respectively; that is, AZU and TZU have similar predictive accuracy, while OLU outperforms them both. Employing a simple neural network technology allows us to obtain more predictive ability out of the Altman, Taffler and Ohlson variables, which is unsurprising given that neural networks are less constrained than MDA and logit. The θ values achieved by our multivariate models in the current setting are, however, lower than the θ values reported in earlier studies: Reisz and Purlich (2007) report θ at 78% for the original Altman model, using US data from 1988-2002; Agarwal and Taffler (2008) report θ at 89% for the original Taffler model, using UK data from 1985-2001; and Duan and Shrestha (2011) report θ at 87.6% for the original Altman model and 88.9% for the original Ohlson model.

Ahead of the multivariate accounting number-based models, the best performing of the models in this study are the contingent claims models. The European call-based model HKCL achieves θ of 82.7%, with its naïve variant, BS, achieving 83.9%. The barrier option-based models achieve similar levels of accuracy, with DOC achieving a θ of 81.8% and its naïve variant, JW_DOC, achieving θ = 83.0%. Our result for the HKCL model is similar to the 84% reported by Agarwal and Taffler (2008); but is much higher than the 71% reported by Reisz and Purlich (2007). As regards model BS, Agarwal and Taffler (2008) report a θ of 87%, which, consistent with our findings, places it above the more sophisticated HKCL.

Of the remaining single variable models, SIZE, with θ = 80.6%, outperforms the multivariate accounting number-based models (indeed, it is in the league of the contingent claims models). Beaver's best performing ratio, BV, at 46 years old, has θ = 67.0% - which places it alongside the multivariate accounting number-based models AZU and TZU in terms of predictive accuracy.

An issue pertaining to the accounting number-based models was the adoption of International Financial Reporting Standards (IFRSs) in the UK from January 2005. Accounting standard changes will certainly impact on the predictive ability of an accounting number-based failure prediction model. As discussed above, the foregoing in this sub-section has reported results based on testing/validation of each of the models using post-IFRS data, but with training/estimation of certain models performed using predominantly pre-IFRS accounting data – the empirical method being designed to: (i) use up-to-date data; (ii) include sufficient failed firms to allow for sound analysis; and (iii) replicate use in practice, by estimating models (where required) on a sample period preceding model application. Avoidance of the transition to IFRS is, for the time being, not realistic and a study based wholly in the post-IFRS era might be an area for future research. In the following sub-section, however, we consider the distributional of properties of θ for each model, derived by a 10,000-iteration Monte Carlo approach to training and validation sample splits – without any chronological constraint on the membership of the training and validation samples, i.e., allowing pre- and post-IFRS data in both training and validation samples.

6.2. Distributional properties of θ

Summary statistics regarding θ for ten[24] of the insolvency prediction models are set out in Table 7; and the relative frequency distributions of θ for each of these models are shown graphically in Fig. 5.

***** insert Table 7 about here *****

and

***** insert Fig. 5 about here *****

Results as to the efficacy of the models as measured by the mean of θ are in agreement with the results presented in section 6.1. The ordering of model efficacy by mean θ from Table 7 is identical to that according to θ from Table 6. Further, each θ from Table 6 is close to the corresponding mean θ from Table 7; indeed, there are no differences between Table 6 and corresponding Table 7 mean θ significant at the 5% level. The results of section 6.1 are not, therefore, outliers or atypical in the context of the distributions of possible θs.

Concerning the ordering of the different models by mean θ, each model's mean θ is different from the mean(s) of the model(s) of neighbouring efficacy with significance at the 1% level. Whilst the means of the distributions ordered by efficacy are significantly different one to the next, it is clear from Fig. 5 that there is considerable overlap in the distributions and that a particular sample split into training and validation samples might suggest an alternative ordering. This leads to discussion of the spread of the distributions.

The less spread in a distribution, the less sensitive the model's efficacy to variation in the composition of training and validation/test samples. Model TZU's θ distribution shows the highest spread, with a standard deviation of 0.0346. Models B/M, AZU, OLU and BV have distributions with the four next highest standard deviations (being 0.0260, 0.0256, 0.0207 and 0.0195 respectively). In this respect, superior again are the contingent claims models, HKCL BS, DOC and JW_DOC, and the single variable SIZE model, with their distributions of θ having standard deviations of 0.0165, 0.0131, 0.0167, 0.0138 and 0.0149 respectively. The relatively high standard deviations of θ distributions for the accounting number-based

[24] The three neural network-implemented models are not included, since the computing time required for 10,000 iterations of the neural network implementations is prohibitive.

multivariate models AZU, TZU and OLU is consistent with, and indeed may explain, the variety of efficiency claims for these models as between different studies.

The θ distributions of all of the models show a reasonable degree of regularity and symmetry but Jarque-Bera testing of the distributions finds that a null of normality may be rejected for seven of the ten models at the 1% level of significance. The exceptions are BV and DOC, where normality may be rejected only at the 5% level of significance and B/M, where normality may not be rejected at any generally acceptable level of significance. It is interesting to note that model B/M, which has the most ‘normal-like’ of the distributions, is by far the least efficient of the models tested.

In sum, our results show that the contingent claims models, in all of the guises we adopt, performed better than the accounting-number based models and that, within the family of accounting number-based models, the single variable approach of Beaver (1966), counter intuitively, can be just as effective as more complex multivariate approaches. Our results are at variance with those of Taffer (2008), who does not find contingent claims models to be superior to the accounting number-based Taffer Z-score model as regards predicting UK insolvencies. Our results are, however, consistent with the US results of Hillegeist et. al (2004) in finding contingent claims models to be superior to accounting number-based models and with the results of the Aziz and Dar (2006) 89-paper international meta-analysis which finds that theoretically-based insolvency prediction models outperform those of a statistical nature and, more specifically, that contingent claims models have higher predictive accuracy than single variable, MDA, logit and NN models.

7. Conclusions

We have presented a review of the methodologies proposed and employed for the prediction of corporate insolvency over the last five decades, and catalogued 25 different methods. Informed by prior frequencies of use and reported efficacies of the various models, we select a variety for testing; and develop and test a new naïve version of the down-and-out call contingent claims model. Our study uses up-to-date UK data, including post-IFRS data, and our sample period encompasses the recent financial crisis.

Using ROC curve analysis, we show that the traditional accounting-based models are outperformed in our setting by theoretically-driven contingent claims methods: of thirteen models tested, the four best performing models are contingent claims models based on European call and barrier options. The new, naïve version of the down-and-out call option contingent claim model which we propose is second only to the Bharath and Shumway (2008) naïve contingent claim model in terms of predictive accuracy and each of these models outperforms its more complex antecedent.

As regards multivariate accounting number-based models, we are able to obtain improved predictive ability from the variables of the Altman (1968), Taffler (1983) and Ohlson (1980) models by applying simple neural network technology but this improvement is not sufficient to challenge the superiority of the contingent claims models.

Of the single variable models, one, firm size, compares well against more sophisticated multivariate accounting number-based approaches and another, Beaver's (1966) best performing ratio, cash flow to total debt, has predictive ability at a similar level to the accounting number-based MDA approaches. The third single variable model, book-to-market ratio, used as focal/benchmark variable in earlier studies, is not found to be useful in our setting for predicting firm failure.

We suggest one reason for the relatively poor performance of accounting number-based models might be the global financial problems which straddled the end of our sample period and created higher levels of financial uncertainty and distress for firms in general, as well as making the task of distinguishing between the failing and non-failing firms exceptionally difficult. Since accounting information is historical and does not reflect the current position even on its date of publication, the accuracy of multivariate accounting number-based models in predicting insolvency is unlikely to be stable over a period of significant and general economic turmoil. Contingent claims models, incorporating up-to-date market information, are likely to fare better.

In assessing the efficacy of insolvency prediction models we must be mindful of two impactors (which work contrary to one another) on model efficacy which will arise when models are applied in practice, but which have rarely been mentioned in the academic literature. The first is that an efficient model should become the victim of its own success and this may, to some extent, have been the fate of the more popular accounting numbers-based models. If it is possible to formulate a prediction model of high accuracy, then efficiency arguments would suggest that the model would come into widespread use to identify failing firms and allow targeted interventions designed to save the firms from demise. Even a partial success rate for such interventions would lead to the apparent reduced efficacy of the prediction model (and, indeed, cause a higher concentration of failures owing to, or correlated with, factors not considered by the model). Contrary to this impactor is another, that a model of questionable efficiency may make a self-fulfilling prophecy of firm failure; the mere prediction of failure might cause failure since the prediction leads to withdrawal of funding and trade credit, the migration of staff and a reduced set of strategic options. The practical application of such models should, therefore, be treated with caution.

Given our findings overall, we suggest that further research upon the development and implementation of contingent claims models is desirable and that insolvency prediction based on accounting numbers alone, and the Altman (1968), Ohlson (1980) and Taffler (1983) models, over a quarter of a century old, now give way to techniques which consider market-based information alongside accounting numbers.

Acknowledgements

We thank David Gwilliam and Mark Tippett and participants at the British Accounting and Finance Association 48th Annual Conference, 2011, the Financial Reporting and Business Communication Research Unit 15th Annual Conference, 2011 and the Gregynog Accounting and Finance Colloquium, 2011 for their helpful comments and suggestions on an earlier version of this paper.

Appendix. Specification of MDA and logit models

Altman (1968) MDA model discriminant function

$$Z = \alpha + \beta_1 \frac{WC}{TA} + \beta_2 \frac{RE}{TA} + \beta_3 \frac{EBIT}{TA} + \beta_4 \frac{V_E}{TL} + \beta_5 \frac{S}{TA}$$

where *WC/TA* is working capital divided by total assets; *RE/TA* is retained earnings divided by total assets; *EBIT/TA* is earnings before interest and taxes divided by total assets; V_E/TL is market value of equity divided by total liabilities; and *S/TA* is sales divided by total assets.

Taffer (1983) MDA model discriminant function

$$Z = \alpha + \beta_1 \frac{PBT}{CL} + \beta_2 \frac{CA}{TL} + \beta_3 \frac{CL}{TA} + \beta_4 NCI$$

where *PBT/CL* is profit before tax divided by current liabilities; *CA/TL* is current assets divided by total liabilities; *CL/TA* is current liabilities divided by total assets; and *NCI* is the 'no-credit interval', being the period a company can continue its operations using its immediate assets if all other forms of short term finance are cut off. More directly, *NCI* is defined as immediate assets (current assets – stock) less current liabilities, divided by daily operating costs excluding depreciation.

Ohlson (1980) logit model

Probability of firm failure $= \frac{1}{1+e^{-Z}}$

where

$$Z = \alpha + \beta_1 SIZE + \beta_2 \frac{TL}{TA} + \beta_3 \frac{WC}{TA} + \beta_4 \frac{CL}{CA} + \beta_5 OENEG + \beta_6 \frac{NI}{TA} + \beta_7 FUTL + \beta_8 INTWO + \beta_9 CHIN$$

and where *SIZE* is the natural logarithm of GDP-deflated total assets; *TL/TA* is total liabilities divided by total assets; *WC/TA* is working capital divided by total assets; *CL/CA* is current liabilities divided by current assets; *OENEG* is a dummy variable equal to one if total liabilities exceed total assets, and zero otherwise; *NI/TA* is net income divided by total assets; *FUTL* is funds from operations (pre-tax income plus depreciation and amortization) divided by total liabilities; *INTWO* is a dummy variable equal to one if net income was negative over previous two years, and zero otherwise; and *CHIN* is the scaled change in net income calculated as $(NI_t-NI_{t-1})/(|NI_t|+|NI_{t-1}|)$, where NI_t is the net income for the most recent period.

References

Abrahams, C. R., & Zhang, M. (2009). A Comprehensive Credit Assessment Framework (SAS white paper), available on the internet at http://www.sas.com/resources/whitepaper/wp_4351.pdf

Agarwal, V., & Taffler, R. J. (2007). Twenty-five years of the Taffler z-score model: does it really have predictive ability? *Accounting and Business Research, 37(4)*, 285-300.

Agarwal, V., & Taffler, R. J. (2008). Comparing the performance of market-based and accounting-based bankruptcy prediction models. *Journal of Banking and Finance, 32(8)*, 1541-1551.

Altman, E. I. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance, 23(4)*, 589-609.

Altman, E. I., & Eisenbeis, R. A. (1978). Financial Applications of Discriminant Analysis: A Clarification. *Journal of Financial and Quantitative Analysis, 13(1)*, 185-195.

Altman, E. I., Haldeman, R. G., & Narayanan, P. (1977). ZETA analysis: A new model to identify bankruptcy risk of corporations. *Journal of Banking and Finance, 1(1)*, 25-54.

Aziz, M., & Dar, H. (2006). Predicting corporate bankruptcy: where we stand. *Corporate Governance, 6(1)*, 18-33.

Balcaen, S., & Ooghe, H. (2006). 35 years of studies on business failure: an overview of the classic statistical methodologies and their related problems. *The British Accounting Review, 38(1)*, 63-93.

Beaver, W. (1966). Financial Ratios as Predictors of Failures. *Journal of Accounting Research, 4(3)(Supplement)*, 71-102.

Beaver, W. (1968). Market prices, financial ratios and prediction of failure. *Journal of Accounting Research, 6(2)*, 179-192.

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

Bharath, S., &. Shumway, T. (2008). Forecasting Default with the Merton Distance to Default Model. *The Review of Financial Studies, 21(3)*, 1339-1369.

Black, F., & Cox, J. (1976). Valuing Corporate Securities: Some Effects of Bond Indenture Provisions, *Journal of Finance, 31(2)*, 351-367.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities, *Journal of Political Economy, 81(3)*, 637-654.

Blum, M. (1974). Failing Company Discriminant Analysis. *Journal of Accounting Research, 12(1)*, 1-25.

Boritz, J. E., Kennedy, D. B., & Albuquerque, A. (1995). Predicting corporate failure using a neural network approach. *Intelligent Systems in Accounting, Finance and Management, 4(2)*, 95-111.

Brockman, P., & Turtle, H. J. (2002). A barrier option framework for corporate security valuation. *Journal of Financial Economics, 67(3)*, 511-529.

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

Coats, P.K., & Fant, L.F. (1993). Recognizing financial distress patterns using a neural network tool. *Financial Management*, 22(3), 142-155.

Cooper, D. (2010). Company insolvencies rise dramatically in 2009 with little hope of improvement for 2010 (CooperMatthews Debt Articles), available on the internet at http://coopermatthews.com/articles/2010/Company-insolvencies-rise-dramatically-in-2009.html.

Crosbie, P. J., & Bohn, J. R. (2003). Modelling Default Risk. *Moody's KMV*.

Deakin, E. (1972). A Discriminant Analysis of Predictors of Business Failure. *Journal of Accounting Research*, 10(1), 167-179.

Deakin, E. (1977). Business Failure Prediction: An Empirical Analysis. In Altman, E. I., & Sametz, A. W. (Eds.), *Financial crises: Institutions and Markets in a Fragile Environment* (pp. 69-88). New York: John Wiley & Sons.

Dhrymes, P. (1974). *Econometrics*, New York: Springer-Verlag.

Diamond, H. (1976). Pattern recognition and the detection of corporate failure (PhD thesis, New York University).

Duan, J., & Shrestha, K. (2011). Statistical Credit Rating Methods', *Global Credit Review, 1*, 43-64.

Edmister, R. (1972). An empirical test of financial ratio analysis for small business failure prediction, *Journal of Financial and Quantitative Analysis, 7(2)*, 1477-1493.

Eisenbeis, R.A. (1977). Pitfalls in the Application of Discriminant Analysis in Business, Finance, and Economics. *Journal of Finance, 32(3)*, 875-900.

Engelmann, B., Hayden, E., & Tasche, D. (2003). Testing rating accuracy. *Risk, 16*, 82-86.

Falkenstein. E., Boral, A., & Carty, L. (2000). RiskCalc for private companies: Moody's default model. *Moody's KMV*.

Geske, R. (1977). The Valuation of Corporate Liabilities as Compound Options. *Journal of Financial and Quantitative Analysis, 12(4)*, 541-552.

Gönen, M. (2006). Receiver operating characteristic curves. *Proceedings of the SAS Users Group International 31 Conference, San Francisco, CA*, available on the internet at http://www2.sas.com/proceedings/sugi31/210-31.pdf.

Gujarati, D. (2003). *Basic Econometrics, Fourth Edition.* London: McGraw-Hill.

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.

Hanley, J., & McNeil, B. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143, 29-36.

Hebb, D. (1949). *Organization of Behavior.* New York: John Wiley.

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

Joy, O.M., & Tollefson, J. O. (1975). On the financial applications of discriminant analysis. *Journal of Financial and Quantitative Analysis, 10(5)*, 723-739.

Joy, O.M., & Tollefson, J. O. (1978). Financial Applications of Discriminant Analysis: A Clarification. *Journal of Financial and Quantitative Analysis, 13(1)*, 185-195.

Keasey, K., & McGuinness, P. (1990). The failure of UK industrial firms for the period 1976-1984, logistical analysis and entropy measures. *Journal of Business Finance & Accounting, 17(1)*, 119-135.

Keasey, K., & Watson, R. (1986). The Prediction of Small Company Failure: Some Behavioural Evidence for the UK. *Accounting and Business Research*, *17(65)*, 49-58.

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

Law, S. (2010). Media Statement: R3 comments on the insolvency statistics for Q1. *R3 Press Releases, 7th May,* available on the internet at http://www.creditman.co.uk/uk/members/news-view.asp?newsviewID=11742.

Libby, R. (1975). Accounting Ratios and the Prediction of Failure: Some Behavioural Evidence, *Journal of Accounting Research*, 13, (1), 150-161.

Maddala, G. (1983). *Econometric Society Monographs: Limited-dependent and qualitative variables in econometrics*. Cambridge: Cambridge University Press.

McCulloch, W., & Pitts, W. (1943). A logical calculus of the ideas imminent in nervous activity. *Bulletin of Mathematical Biophysics*, *5*,119-133.

McDonald, R. (2002). *Derivative Markets*. Boston, MA: Addison Wesley.

McFadden, D. (1973). Conditional Logit Analysis of Qualitative Choice Behaviour. In P. Zarembka (Ed.), *Frontiers in Econometrics* (pp. 105-142). New York: Academic Press.

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance, 29(2)*, 449-470.

Moyer, R. C. (1977). Forecasting financial failure: a re-examination. *Financial Management, 6(1)*, 11-17.

Odom, M., & Sharda, R. (1990). A neural network model for bankruptcy prediction. *Proceedings of the IEEE International Joint Conference on Neural Networks, San Diego, CA,* II.63-II.67.

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

Ooghe, H., Joos, P., De Vos, D., & Bourdeaudhuij, C. (1994). Towards an improved method of evaluation of financial distress models and presentation of their results (Research report, Ghent University).

R3 (2010). The 'insolvency lag': risks for 2010: Evidence from previous recessions shows early recovery is a dangerous time for struggling businesses and individuals (R3 Briefing Paper), available in the internet at http://www.r3.org.uk/media/documents/policy/policy_papers/corporate_insolvency/The_Insolvency_Lag_risks_for_2010.pdf.

Reisz, A., & Purlich, C. (2007). A market-based framework for bankruptcy prediction. *Journal of Financial Stability, 3(2),* 85-131.

Rojas, R. (1996). *Neural Networks*, Berlin: Springer-Verlag.

Rosenblat, F. (1957). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review, 65*, 386-408.

Scott, J. (1981). The Probability of Bankruptcy: A Comparison of Empirical Predictions and Theoretical Models. *Journal of Banking and Finance, 5(3)*, 317-344.

Shumway, T. (1999). Forecasting bankruptcy more accurately. A simple hazard model (Working paper, University of Michigan), .available on the internet at http://textbiz.org/projects/defaultprediction/hazardmodel.pdf.

Sobehart J., & Stein, R. (2000). Moody's Public Firm Risk Model: A Hybrid Approach To Modeling Short Term Default Risk. *Moody's KMV.*

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

Taffler, R. J. (1983). The Assessment of Company Solvency and Performance Using a Statistical Model. *Accounting and Business Research, 13(52)*, 295-308.

Taffler, R. J., & Tisshaw, H. (1977). Going, Going, Gone - Four Factors Which Predict. *Accountancy, 88(1003)*, 50-54.

Tudela, M., & Young, G. (2003). A Merton-Model Approach to Assessing the Default Risk of UK Public Companies (Working paper, Bank of England), available on the internet at http://www.bankofengland.co.uk/publications/Documents/workingpapers/wp194.pdf.

Vassalou, M., Xing, Y. (2004). Default Risk in Equity Returns. *Journal of Finance, 59(2)*, 831-868.

Zavgren, C.V. (1983). The prediction of corporate failure: the state of the art. *Journal of Accounting Literature, 2*, 1-38.

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

Zmijewski, M. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research, 22 (Supplement),* 59-86.

Figures and tables

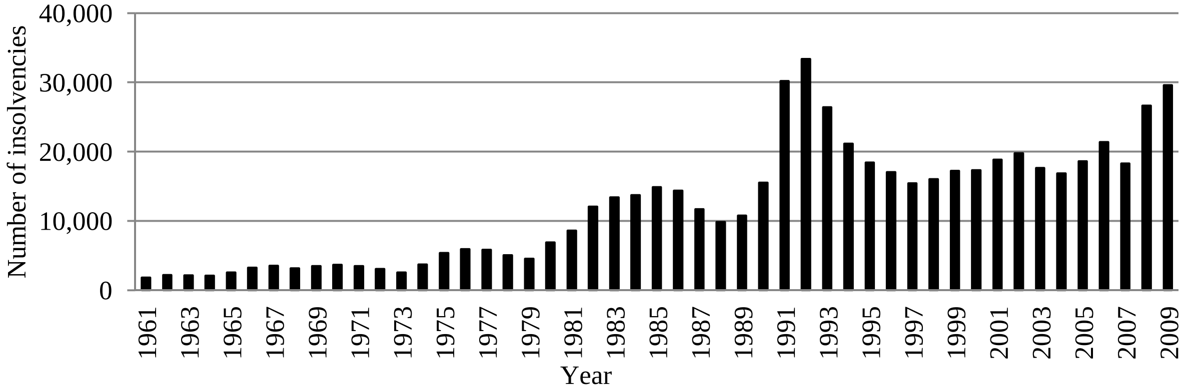

Fig. 1. Total number of UK insolvencies by year, 1961-2009. Data source: www.insolvency.gov.uk

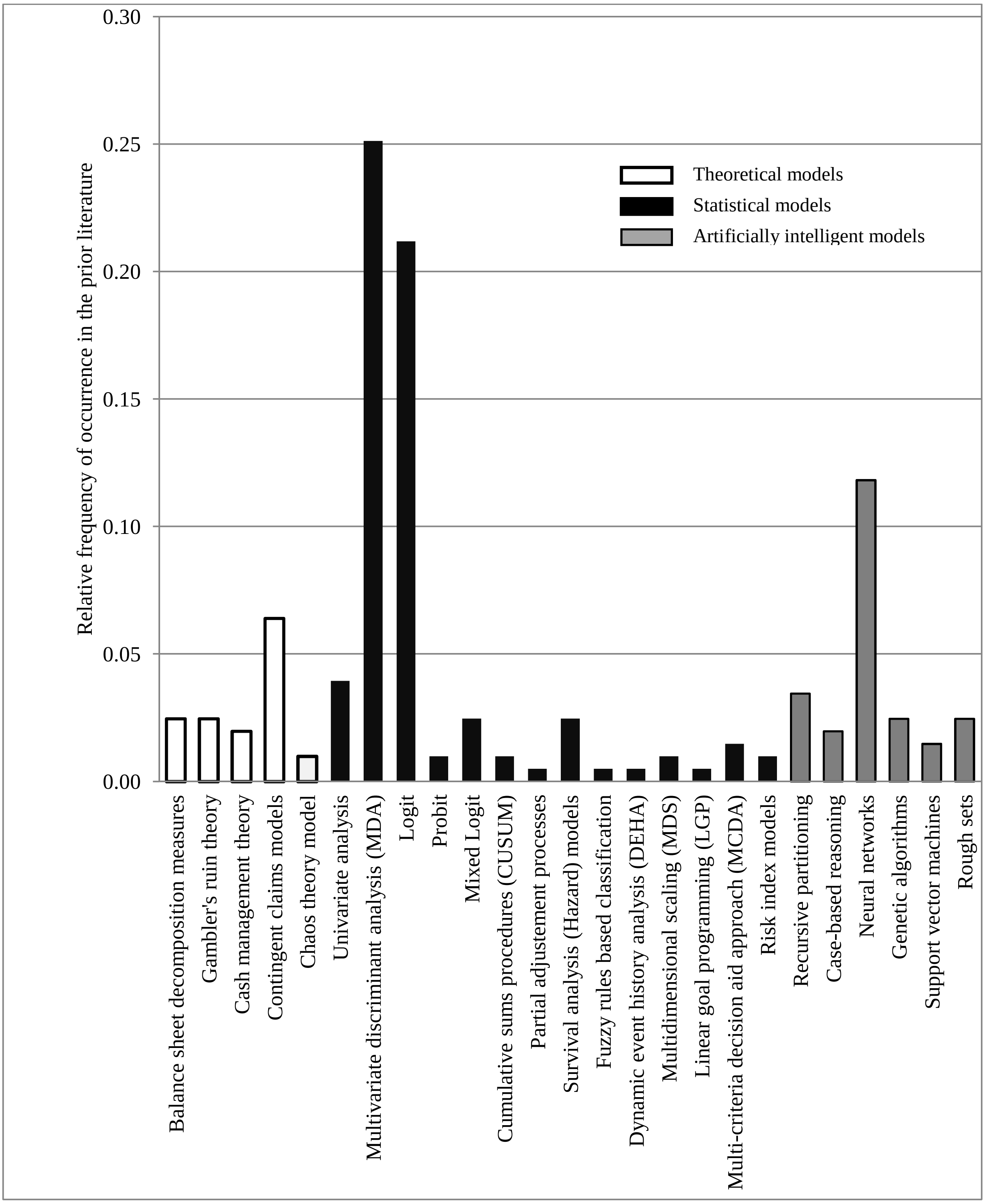

Fig. 2. Relative frequency of occurrence of various insolvency prediction techniques in the prior literature.

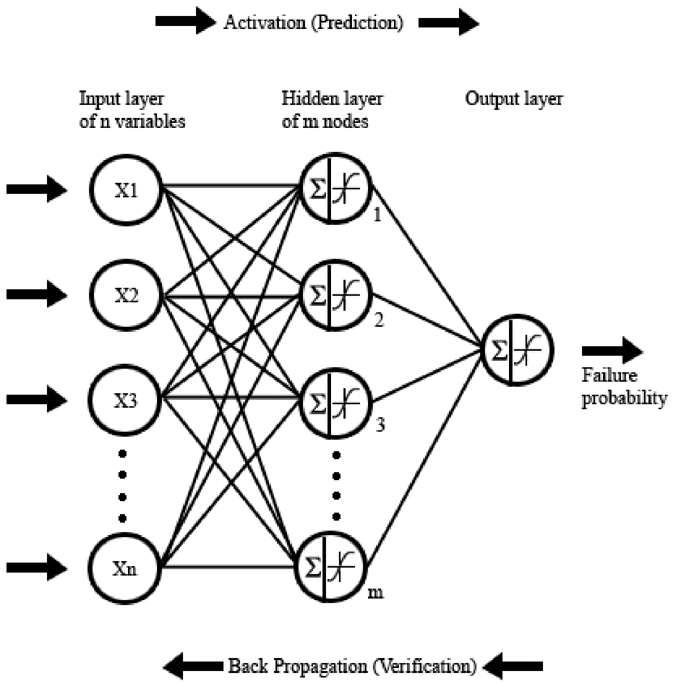

Fig. 3. Outline schematic for a feed-forward neural network with back propagation.

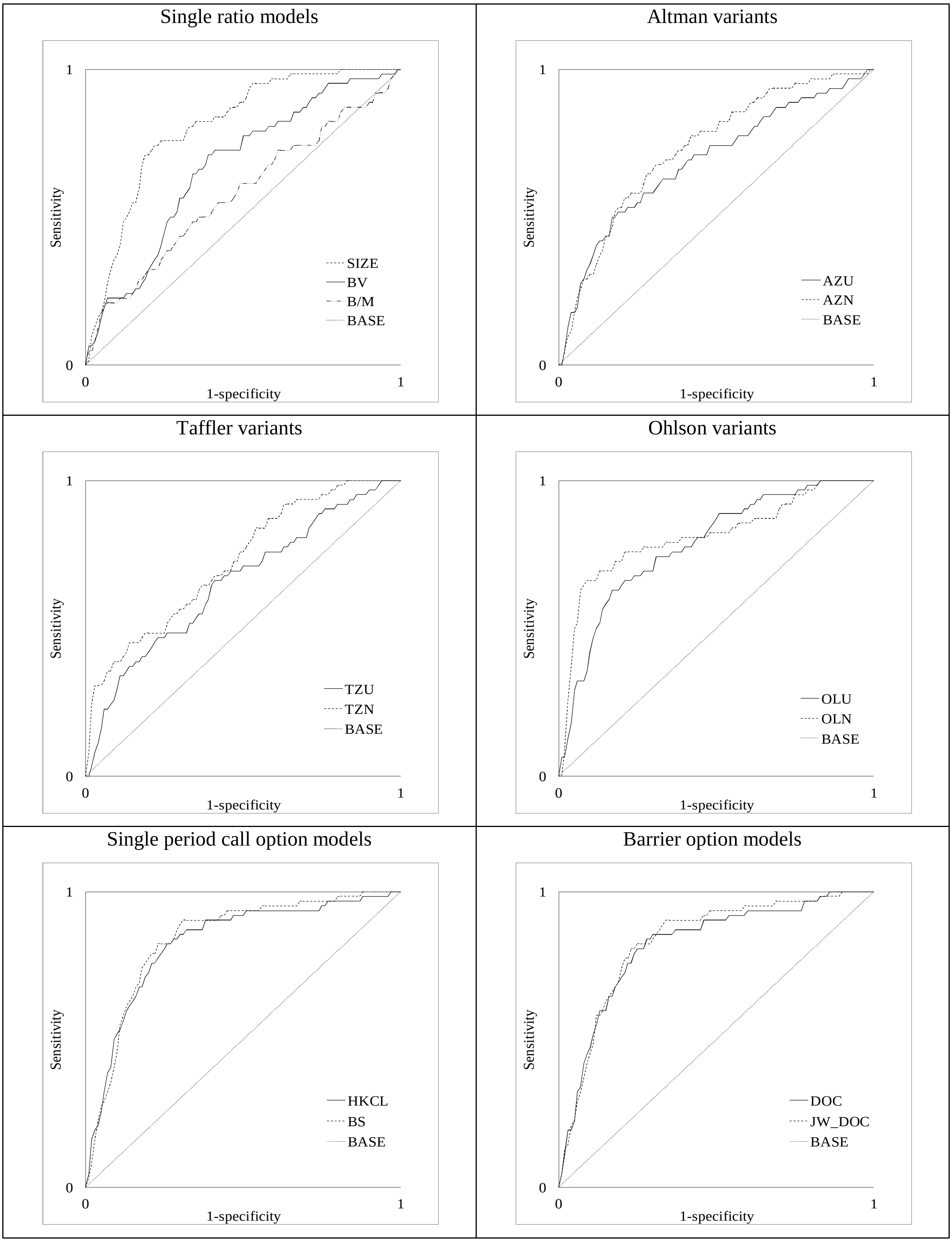

Fig. 4. Receiver operating characteristic (ROC) curves. Note: the 45-degree line (BASE) in each case represents a model with no predictive accuracy (i.e., a random choice model).

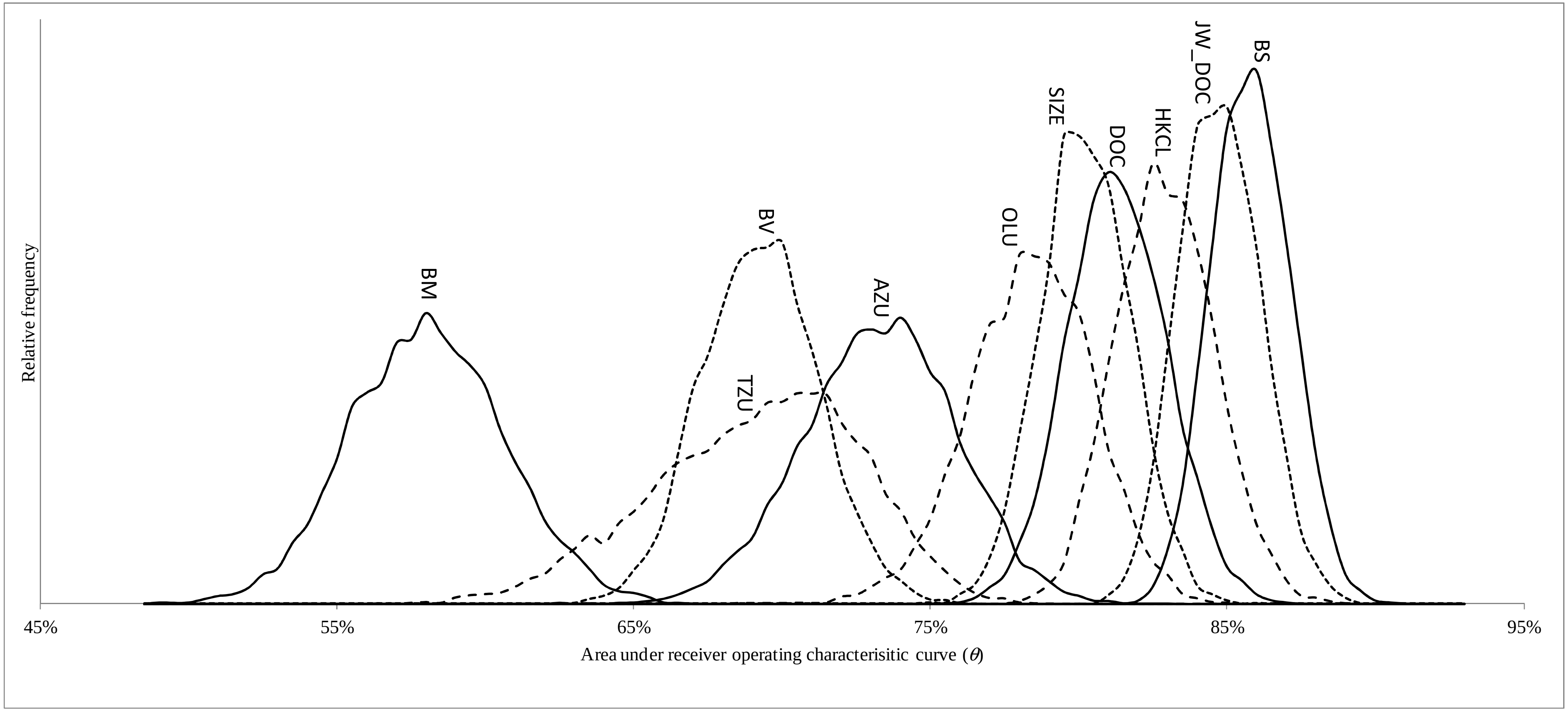

Fig. 5. Relative frequency distributions of area under receiver operating characteristic curves (θ) for ten insolvency prediction models. Based on 10,000 random splits between training and validation samples.

Table 1
Summary of the sample sizes analysed in a selection of prior studies.

Author(s) (year)	Primary model type	Country	Failed	Non-failed
Beaver (1966)	Single Ratio	US	79	79
Altman (1968)	MDA	US	33	33
Ohlson (1980)	Logit	US	105	2,000
Taffler (1983)	MDA	UK	46	46
Sobehart and Stein (2000)[a]	Options	US	1,406	13,041
Charitou et al. (2004)	Neural network	UK	51	51
Bharath and Shumway (2008)	Options	US	1,449	Not disclosed
Hillegeist et al. (2004)	Options	US	756	14,303
Reisz and Purlich (2007)	Options	US	799	4,985
Agarwal and Taffler (2008)	MDA/Options	UK	103	2,006

[a] Moody's RiskCalc model for public companies

Table 2
The number of firm year observations employed in this study.

	Training (2000-2005)	Validation (2006-2009)	Total (2000-2009)
Failed	39	62	101
Non-Failed	2,964	3,530	6,494
Total	3,003	3,592	6,595

Table 3
Models selected for testing: key characteristics.

Model [model designation]	Data	Implementation	Type[a]
SIZE	Market	Single variable	Statistical
Book to Market [B/M]	Accounting	Single variable	Statistical
Beaver (1966) best performing ratio CF/TD [BV]	Accounting	Single variable	Statistical
Altman Z-score with updated coefficients [AZU]	Accounting	Multivariate, MDA	Statistical
Altman Z-score variables, implemented using NN [AZN]	Accounting	Multivariate, NN	AIES
Taffler UK Z-score with updated coefficients [TZU]	Accounting	Multivariate, MDA	Statistical
Taffler UK Z-score variables, implemented using NN [TZN]	Accounting	Multivariate, NN	AIES
Ohlson logit model with updated coefficients [OLU]	Accounting	Multivariate, logit	Statistical
Ohlson logit model variables, implemented using NN [OLN]	Accounting	Multivariate, NN	AIES
Hillegeist et al. (2004) market model [HKCL]	Market	Contingent claims	Theoretical
Bharath and Shumway (2008) naïve market model [BS]	Market	Contingent claims	Theoretical
Barrier option as down-and-out call [DOC]	Market	Contingent claims	Theoretical
New naïve down-and-out call [JW_DOC]	Market	Contingent claims	Theoretical

[a] Type by reference to Aziz and Dar (2006); 'AIES' represents artificially intelligent expert systems.

Table 4
Classical form of summary tabulation of model prediction versus actual outcome data (number of firms) for the purposes of efficiency assessment.

	Actual outcome for firm		
	Failed	Non-failed	Total
Predicted outcome:			
Failed	*TP*	*FP*	*TP*+*FP*
Non-failed	*FN*	*TN*	*FN*+*TN*
Total	*TP*+*FN*	*TN*+*FP*	*TP*+*FP*+*TN*+*FN*

TP (True Positive) is the correct classification of a failed firm; and *TN* (True Negative) is the correct classification of a non-failed firm. A false negative outcome (*FN*) is predicting a firm to survive when it actually fails (type I error); and a false positive outcome (*FP*) is predicting a firm to fail when it actually survives (type II error).

Table 5
Mean predicted probabilities of failure for failed versus non-failed groups of firms.

Model [model designation]	All firms	Non-failed firms	Failed firms	*t*-stat[a]
Single variable approaches				
SIZE	0.066	0.063	0.198	6.400**
Book to Market [B/M]	0.738	0.737	0.781	2.385**
Beaver (1966) best performing ratio CF/TD [BV]	0.307	0.305	0.404	3.734**
Altman Z-score with updated coefficients [AZU]	0.401	0.399	0.520	5.334**
Altman Z-score variables, implemented using NN [AZN] [b]	0.012	0.012	0.012	0.003
Taffler UK Z-score with updated coefficients [TZU]	0.433	0.431	0.497	3.937**
Taffler UK Z-score variables, implemented using NN [TZN] [b]	0.032	0.031	0.034	0.171
Ohlson logit model with updated coefficients [OLU]	0.015	0.014	0.040	1.500
Ohlson logit model variables, implemented using NN [OLN] [b]	0.013	0.013	0.044	1.830*
Hillegeist et al. (2004) market model [HKCL]	0.011	0.009	0.067	3.317**
Bharath and Shumway (2008) naïve market model [BS]	0.024	0.022	0.115	5.019**
Barrier option as down-and-out call [DOC]	0.012	0.011	0.069	3.334**
New naïve down-and-out call [JW_DOC]	0.047	0.044	0.221	7.800**

**, * denote one-tailed significance at the 1% level and 5% level respectively
[a] Test of difference in mean predicted probability of failure, non-failed firms versus failed firms.
[b] The relative magnitude of the neural network estimates are not directly comparable to the other models, since the validating scores obtained from the hidden network layer tended to be very small and the groups are separable only at four to five decimal places.

Table 6
Efficiency of tested models: area under receiver operating characteristic curves (θ) and accuracy ratio (*AR*).

Model [model designation]	θ	SE_θ	Z	AR
Single variable approaches				
SIZE	80.6%	0.034	9.097**	0.61
Book to Market [B/M]	57.7%	0.034	2.004*	0.15
Beaver (1966) best performing ratio CF/TD [BV]	67.0%	0.038	4.461**	0.34
Altman Z-score with updated coefficients [AZU]	68.9%	0.038	4.994**	0.38
Altman Z-score variables, implemented using NN [AZN]	73.1%	0.037	6.293**	0.46
Taffler UK Z-score with updated coefficients [TZU]	65.3%	0.038	3.992**	0.31
Taffler UK Z-score variables, implemented using NN [TZN]	72.4%	0.037	6.046**	0.45
Ohlson logit model with updated coefficients [OLU]	78.1%	0.035	8.048**	0.56
Ohlson logit model variables, implemented using NN [OLN]	80.9%	0.033	9.242**	0.62
Hillegeist et al. (2004) market model [HKCL]	82.7%	0.032	10.076**	0.65
Bharath and Shumway (2008) naïve market model [BS]	83.9%	0.032	10.733**	0.68
Barrier option as down-and-out call [DOC]	81.8%	0.033	9.655**	0.64
New naïve down-and-out call [JW_DOC]	83.0%	0.032	10.267**	0.66

Area under the receiver operating characteristic curve, θ, represents a model's overall predictive accuracy. It is calculated using the trapezoid method, which Hanley and McNeil (1982) show to be equivalent to the Wilcoxon test statistic. The standard error of θ, SE_θ, is also calculated following Hanley and McNeil (1982). *AR*, following Englemann et al. (2003), is a linear transformation of θ: $AR = 2(\theta - 0.5)$. The *Z* column shows the test statistic on deviation from a null hypothesis of $\theta = 50\%$, i.e., no predictive ability in excess of that of a random assignment between failed and non-failed.
**, * denote one-tailed significance at the 1% level and 5% level respectively

Table 7
Distributional properties[a] of model predictive accuracy, as measure by area under receiver operating characteristic curves (θ).

Model [model designation]	Mean	Median	SD	Skewness	Excess kurtosis[b]	Minimum	Maximum	Jarque-Bera ~ $\chi^2(2)$
Single variable approaches								
SIZE	80.5%	80.5%	0.0149	0.0403	-0.1327	75.1%	86.3%	10.04**
Book to Market [B/M]	58.3%	58.3%	0.0260	0.0391	-0.0314	49.1%	68.8%	2.96
Beaver (1966) best performing ratio CF/TD [BV]	69.5%	69.5%	0.0195	0.0472	-0.0823	62.9%	76.7%	6.53*
Altman Z-score with updated coefficients [AZU]	73.5%	73.6%	0.0256	-0.1206	0.0979	60.4%	83.0%	28.25**
Taffler UK Z-score with updated coefficients [TZU]	69.4%	69.7%	0.0346	-0.3681	-0.1861	54.2%	79.3%	240.24**
Ohlson logit model with updated coefficients [OLU]	78.6%	78.7%	0.0207	-0.2298	0.2614	68.6%	87.1%	116.51**
Hillegeist et al. (2004) market model [HKCL]	83.2%	83.1%	0.0165	0.1116	-0.1266	77.8%	89.0%	27.43**
Bharath and Shumway (2008) naïve market model [BS]	86.1%	86.0%	0.0131	0.1103	-0.1995	81.9%	91.0%	36.84**
Barrier option as down-and-out call [DOC]	81.5%	81.5%	0.0167	0.0648	-0.0605	75.4%	87.8%	8.52*
New naïve down-and-out call [JW_DOC]	85.0%	85.0%	0.0138	0.1319	-0.1511	80.5%	89.9%	38.52**

[a] Based on 10,000 random splits between training and validation samples. In each case the number of both failed firm-year observations and non-failed firm year observations in the training and validation samples are as in Table 2; but the allocations now being made on a random basis, rather than according to chronology.
[b] Excess kurtosis is the unbiased measure of deviation of kurtosis from 3 (3 being the kurtosis of a normal distribution).
**, * denote one-tailed significance at the 1% level and 5% level respectively