Discovering the optimal set of ratios to use in accounting-based models

Duarte Trigueiros
ISTAR, University Institute of Lisbon
duarte.trigueiros@iscte-iul.pt
and
Carolina Sam
European Studies Alumni Association, Macau SAR, China
carolsam@al.gov.mo

July 7, 2019

Abstract

Ratios are the prime tool of financial analysis. In predictive modelling tasks, however, the use of ratios raises difficulties, the most obvious being that, in a multivariate setting, there is no guarantee that the collection of ratios eventually selected as predictors will be optimal in any sense. Using, as starting-point, a formal characterization of cross-sectional accounting numbers, the paper shows how the Multilayer Perceptron can be trained to create internal representations which are an optimal set of ratios for a given modelling task. Experiments suggest that, when such ratios are utilised as predictors in well-known modelling tasks, performance improves on that reported by the extant literature.

KEYWORDS: Multilayer Perceptron, Knowledge extraction, Financial analysis, Financial ratio, Sirm size, Bankruptcy prediction, Financial misstatement detection, Earnings forecast.

JEL CODES: C45, G33, M41

REFERENCING: Trigueiros, D., Sam, C. (2018), Discovering the optimal set of ratios to use in accounting-based models. International Journal of Society Systems Science, 10 (2), 110-131.

1 Introduction

At their inception, Neural Networks were supposed to be capable of highlighting features that might be implicit in data. The Multilayer Perceptron (MLP), for instance, was expected to form meaningful "internal representations" in its hidden nodes, so that features of a given relationship would become visible (Rumelhart et al., 1986). Further, it was hoped that such features, once brought to light, would shed light on the type of mechanisms at work in the relationship. In short, Neural Networks were viewed, not just as versatile and powerful modelling machines, but also as potential knowledge-extraction tools. With the passing of the years, such hopes proved to be somehow elusive, as instances of knowledge-extraction via MLP's were not many. Plausibly, knowledge-extraction seems to be feasible in cases where knowledge is well-defined, but less so where knowledge is not clearly circumscribed.

The paper describes an instance of circumscribed knowledge where the promise of knowledge-extraction via MLP's internal representations is fulfilled. When a cross-section of accounting numbers is set to predict an attribute such as imminent bankruptcy, a specifically trained MLP is able to find sets of ratios with optimal predictive power. Such set can then be used to model the relationship. In this way, the assistance of a financial analyst, which would otherwise be required to select predictors, can be dispensed with. Indeed, financial analysts may benefit from knowing the set of ratios which is, in a multivariate context, suitable for a given predictive task.

Findings reported in the paper are timely as financial institutions, regulators and individual investors, increasingly rely on accounting-based models to assess credit risk, to estimate the likelihood of default, to forecast Earnings' changes or the growth of dividends, to detect financial misstatement and to carry out other predictive tasks (Amani and Fadlalla, 2017).

The use of ratios as predictors in multivariate models seems to be an extrapolation of their early use in financial analysis. Since long ago, ratios are a major tool in the analysis of financial statements where they highlight financial features such as profitability and liquidity, while allowing the comparison of varied-size companies. But the fact that ratios are suited to financial analytic tasks does not grant that they may be equally suited in predictive modelling tasks. The need to select an appropriate set of ratios and the varied distributions of ratios are just the most glaring difficulties facing those who want to build accounting-based predictive models using ratios.

The methodology presented in the paper is meant to overcome such difficulties. First, a formal basis for ratio usage is applied to the modelling of relationships involving accounting data. Guidelines ensuing from such basis are then adopted in the selection and transformation of MLP inputs as well as in MLP training, so that a set of ratios possessing optimal predicting power is formed in MLP hidden nodes. Finally, application examples are presented, showing that the methodology is feasible.

The remaining of this introduction briefly reviews the relevant literature; then, financial analytic requirements are compared to multivariate modelling requirements with emphasis on the mentioned difficulties.

1.1 Literature review

Neural networks have been extensively applied to accounting and finance research and practice (Trippi and Turban, 1992). Major application areas are the analysis of the financial condition of companies (Landajo et al., 2007), the detection of financial statement fraud (Ngai et al., 2011), the prediction of bankruptcy (Peat and Jones, 2012), the assessment of credit risk and credit rating (Hong and Shin, 2007), the assessment of consumer creditworthiness (Sustersic et al., 2009), the forecasting of earnings (Cao and Parry, 2009), and the forecasting of market trends and risk (McNelis, 2004). An increasing demand for "Fin-Tech" applications may have fostered the development of undisclosed algorithms where Neural Networks may play a role.

Most of the mentioned applications use neural networks simply as predicting machines, with little regard for their knowledge-extraction capabilities. Exceptions (Cinca, 1998) do not abound, at least in this specific application area. The present paper fulfils a promising intuition put forward by previous research (Berry and Trigueiros, 1993) where it was suggested that, given the approximately Lognormal distributions of line items taken from cross-sections of sets of accounts, nodes in the hidden layer of an MLP would form internal representations which would be, in logarithmic space, quotients of several such items; and it was hoped that such quotients would convey financial-analytic meaning.

1.2 Ratios as financial-analytic tools

Business companies, namely those listed in stock markets, are required to regularly give a detailed record of their financial position and output: revenues and expenses incurred in during the period, the value of assets and liabilities at the end of the period, owners' position, cash-flows and others. Such reports are obtained via a bookkeeping process involving recognition, adjustments and aggregation into specific "line items", of all meaningful transactions of the period. The ensuing "set of accounts" is made available to the public together with notes and auxiliary information, being known as the "financial statement" of the company for that period.

Once published, financial statements are scrutinised by investors, banks and other entities, with the object of supporting decisions concerning companies or industrial sectors. Such scrutiny, and the corresponding diagnostic, is known as "financial analysis". The analysis of a company's report is based on comparisons with industry norms or with other companies' reports, and with the same company's report in previous periods. The major tool used by analysts to perform such comparisons is the "ratio", a quotient of appropriately chosen line items, typically taken from the same set of accounts. Conveniently chosen ratios are able to assess financial features which are implicit in reported numbers: liquidity, solvency, profitability and other features. For instance, when a company's income at the end of a given period is compared with the assets required to generate such income, an important feature, profitability, emerges. Since the effect of company size is similar in line items taken from the same company and period, size cancels out when a ratio is formed. Thus, by using ratios, analysts are able to compare features of companies of different sizes. In turn, financial features may convey a clear picture of a company's future economic prospects.

1.3 The selection of ratios to use in multivariate models

The future state of vital attributes such as Earnings' changes (upwards or downwards), solvency (solvent or bankrupt), trustworthiness (fair or misstated set of accounts), can also be predicted, with varying success, from financial statements of companies; and, as mentioned, multivariate modelling algorithms have been extensively employed (Amani and Fadlalla, 2017) to improve on the accuracy with which such states are predicted via financial analysis.

Models where accounting information is required invariably use financial ratios as predictors. This seems to be a natural extension of the essential use of ratios in financial analysis. Such extension is uncontentious–it was neither disputed nor validated–but it may entail difficulties. To begin with, it is somehow contradictory to employ, as multivariate modelling predictors, the type of ratio commonly used by analysts. Modelling algorithms are synthesisers. They build multivariate distributions (moments and covariance matrices) and then they extract from such distributions, not from any specific predictor, whatever variability is optimal in modelling a relationship. Analysts, by contrast, often rely on narrowly-defined ratios: they assess one piece of information at a time, separating knowledge in order to rearrange it in an insightful way. The consequence of transposing to the synthetic domain an essentially analytical routine is that algorithms where ratios are used as predictors may become limited in their choices, unable to use all information potentially available in the corresponding sets of accounts.

It is desirable that multivariate predictors convey general, not overly circumscribed variability. The methodology proposed in the paper employs line items, not predefined ratios, as the starting-point of accounting-based predictive tasks. Then, at a second stage, an MLP finds, based on previously selected items, the set of ratios capable of conveying such variability.

The alternative to the proposed methodology would be to use trial and error to find the optimal set of ratios. But since most ratios are built from 50 line items (1225 different ratios) or more, trial and error is not workable. Undoubtedly, the intervention of an analyst would reduce the number of ratios to be tested; but then, as mentioned, analytic knowledge would not be appropriate here. Indeed, even when performing the most straightforward predictive tasks, authors avoid analytically-wise ratios (Dechow et al., 2011), favouring the use of input-selection algorithms to choose ratios with explanatory power from a large, but far from comprehensive set of ratios.

1.4 The peculiar cross-section distribution of accounting numbers

Distributions of financial statement numbers, as well as their ratios, tend to obey a multiplicative law of probabilities, not an additive law (McLeay and Trigueiros, 2002). It is since long established that distributions of income, wealth, firm size and other economic accruals, obey multiplicative random processes such as the Pareto or the Gibrat laws, not additive processes such as the Central Limit theorem (Ijiri and Simon, 1964). Disagreement exists on the exact shape of such distributions (Ijiri and Simon, 1974), which approximate the Gibrat law for lower- and middle-sized numbers (Cabral and Mata, 2003) and the Pareto law for higher sizes (Angelini and Generale, 2008) but it is widely accepted that economic accruals stem from the accumulation of random amounts, not from expected random effects.

The difference between additive and multiplicative distributions is huge. In additive data it is rare to observe cases which are more than three standard deviations above or below the mean. The height of human adults, for instance, is limited: humans are neither taller than mountains nor smaller than mice. Yet, for multiplicative data, such disproportions are, not just possible but likely: within the same cross-sectional sample, one firm may be 500 times bigger than another.

Difficulties posed by the multiplicative character of ratios have prompted authors to propose ad-hoc remedies, namely the pruning of outliers, ratio transformation, or both (Nikkinen and Sahlström, 2004). But where the multiplicative character of financial statement data is ignored, any subsequent effort to model such data is fruitless due to the distorting effect of heteroscedasticity and influential cases. Pruning, for instance, is suitable for additive distributions with outliers but is inadequate in the case of multiplicative distributions where extreme observations are not necessarily aberrant (McLeay and Trigueiros, 2002).

The variety of distributions found in ratios must be faced rather than ignored. Recent improved results (Altman and Sabato, 2007) stem from the implicit assumption that ratios are multiplicative. The methodology proposed in the paper suitably addresses the multiplicative character of accounting numbers.

2 Formal basis for ratio usage in predictive modelling

The formal basis of the proposed methodology is now described. The premise is that, safeguarding well-known cases, ratios can be used validly in financial analysis. More than one hundred years of applied use testify in favour of such premise.

Valid use of financial ratios rests on the assumption that ratios are capable of removing the effect of size, thus making financial features comparable across firms of different sizes. In turn, the removal of size requires that ratio components be proportional. From this, two consequences follow. First, a common statistical effect, capable of being removed when a ratio is formed, must be present in components of ratios. Without such common effect, components would be independent and proportionality would not be possible. In addition, such common effect should be capable of explaining differences in size observed among sets of accounts. Second, numbers reported in sets of accounts should have multiplicative rather than additive distributions. Indeed, ratios fail to remove the effect of size unless such effect is multiplicative. There is an intrinsic agreement between ratios and multiplicative distributions: divisions remove common multiplicative effects in the same way subtractions remove common additive effects. Further elaboration on the above reasoning and a review of the literature on the validity of ratios is available (McLeay and Trigueiros, 2002).

The most common instance of multiplicative behaviour is the Lognormal distribution. When lognormality is hypothesised, proportionality requires that the cross-sectional variability of x_{ij}, the i^{th} line item from the j^{th} set of accounts, be described as

$$\log x_{ij} = \mu_i + s_j + \varepsilon_{ij} \tag{1}$$

where the logarithm of x_{ij} is explained as an expectation μ_i containing the overall mean and the deviation from the mean specific to item i, plus s_j, the size effect, specific to set of accounts j, plus size-unrelated residual variability ε_{ij}.

Formulation (1) is not meant to minutely describe the cross-sectional behaviour of line items, but to adequately and plausibly account for the common effect of size. The ε_{ij} are hypothesised to be Normal and size-unrelated but not necessarily independent: after the common effect s_j is accounted for, there may remain positive or negative correlations between any two line items.

2.1 Financial Analysis

Trivially, (1) underlies and endorses the use of ratios. Given two items $i = 1$ and $i = 2$ (Revenue and Expenses for instance) and the corresponding reported amounts x_1 and x_2 from the same set, the logarithm of the ratio of x_2 to x_1 is

$$\log \frac{x_2}{x_1} = (\mu_2 - \mu_1) + (\varepsilon_2 - \varepsilon_1) \tag{2}$$

where the effect of size is no longer present. The median ratio $\exp(\mu_2 - \mu_1)$ is a suitable norm against which comparisons may be made. $\exp(\varepsilon_2 - \varepsilon_1)$ is the deviation from such norm observed in j. For example, $\exp(\varepsilon_{\text{Expenses}} - \varepsilon_{\text{Sales}})$ shows whether the Operating Ratio is above or below an industry norm. Thus $\varepsilon_2 - \varepsilon_1$ is, in logarithms, the information conveyed by a ratio about a financial feature.

To the extent that ratios are suitable as analytical tools, premises leading to (1) should be approximately verified. If this were not so, then ratios from large firms would behave differently from ratios from small firms. Profitability, for instance, would not be of any use in cross-section analysis.

Formulation (1) cannot be utilised to explain negative values, such as those found in Net Income. Line items where negative values occur are subtractions of two positive-only items. Net Income, for instance, is the subtraction of Total Costs from Total Revenue. Consider $x = x_A - x_B$ where A and B are positive-valued items. For any sign of x, an algebraic manipulation, intended to compute the logarithm of a subtraction when the logarithms of numbers being subtracted are known, leads to

$$\log|x_A - x_B| = \log x_A + \log|1 - \frac{x_B}{x_A}| \tag{3}$$

Since x_A is positive-valued, (1) applies to $\log x_A$. And since, as shown in (2), x_B/x_A is size-independent, s_j is not present in $\log|1 - x_B/x_A|$. Using the notation adopted in (1), the expected value of $\log|x_A - x_B|$ is written as μ', being the result of adding μ_A to the expected value of $\log|1-x_B/x_A|$. Residuals are similarly written as ε'. From (3),

$$\log|x| = \mu' + s_j + \varepsilon' \tag{4}$$

for positive- and negative-valued x. Although μ' and ε' in (4) no longer have the meaning of μ and ε in (1), notably s_j in (4) is exactly the same as in (1). This is the reason why ratios succeed in removing the effect of size, even where one of the components is negative. Now, if ratios can remove size from negative values, it should be possible to use any type of item, not just positive-valued items, in tasks requiring the removal of size. Such goal is accomplished *inter alia* by transforming x into

$$\text{sign of } x \log|x| \tag{5}$$

where the variability of $\log x$ is separately explained as

$$\begin{array}{lll} \log x & = \mu + s_j + \varepsilon & \text{for } x > 1 \text{ and} \\ -\log|x| & = -|\mu + s_j + \varepsilon| & \text{for } x < -1. \end{array}$$

Size, s_j, is the same as in (1). Transformation (5) can be intuitively represented as

$$x \rightarrow \begin{cases} \log x & \text{for } x > 1 \\ -\log|x| & \text{for } x < -1, \end{cases}$$

being known as the Logmodulus (John and Draper, 1980). Within the range $-1 < x < 1$, numbers transformed according to (5) are equivocal, but monotonic alternatives exist.[1] For positive-only items, (5) is the logarithmic transformation. Zero-valued cases should not be transformed. Indeed, zero-valued numbers are not multiplicative because, in order to be formed, multiplicative numbers require an accumulation.

[1] sign of $x \log(|x| + 1)$ and asinh $(x/2)/\log(n)$ for base n logarithms.

2.2 Multivariate modelling

Consider the formulation

$$y = a + b_1 \log x_1 + b_2 \log x_2 + \cdots \tag{6}$$

where $x_1, x_2 \cdots$ are line items taken from the same set of accounts, $b_1, b_2 \cdots$ are model coefficients, and y is the attribute to be predicted. Henceforward, j is assumed. In the case of a binary classifier, y is supposed to be explained by a linear score as, indeed, it is the case in most classifiers. Where items obey (1), then (6) can be written as

$$y = A + b_1\varepsilon_1 + b_2\varepsilon_2 + \cdots + (b_1 + b_2 + \cdots)s \tag{7}$$

where $A = a+b_1\mu_1+b_2\mu_2+\cdots$ is a constant. In (7), variability made available to model y contains two terms: a size-independent term $b_1\varepsilon_1 + b_2\varepsilon_2 + \cdots$, and a size-related term $(b_1 + b_2 + \cdots)s$.

Consider the case where y is size-independent. Coefficients $b_1, b_2 \cdots$ must add to zero in this case, so as to bar size from entering the relationship. And when such size-independent y is predicted by a linear combination $a + b_1 \log x_1 + b_2 \log x_2$ of just two logarithmic-transformed items, then $b_1+b_2 = 0$ or $b_2 = -b_1 = b$ and (7) becomes

$$y = a + b[(\mu_2 - \mu_1) + (\varepsilon_2 - \varepsilon_1)] \text{ or, from (2),}$$

$$y = a + b \log \frac{x_2}{x_1} \tag{8}$$

Ratio x_2/x_1 is formed in response to y's size-independence. Suppose, for instance, that ratio x_2/x_1 predicts bankruptcy accurately. When the logarithms of its two components are shown to a modelling algorithm, the linear combination $a + b_1 \log x_1 + b_2 \log x_2$ will be the best at explaining bankruptcy when $b_2 = -b_1$ because, at that point, the ratio is formed, its predicting power is released, size is removed. Contingent on the role x_1 and x_2 may play as "numerator" or "denominator" of the ratio in (8), positive or negative b may arise.

Similarly, where size-independent y is predicted by three logarithmic-transformed items, then $b_3 = -b_1 - b_2$ and the equivalent to (8) now is

$$y = a + b_1 \log \frac{x_1}{x_3} + b_2 \log \frac{x_2}{x_3}.$$

Again, in response to size-independence, two ratios are formed. Alternative combinations are possible, signs of coefficients changing accordingly.

In the general case, size-independent relationships might indeed induce $N-1$ ratios from N line items; but there is no reason why the algorithm, unless required to do so, should pair items two-by-two. Instead, size-related variability will be allotted to a number of items in order to offset size-related variability in other items. The algorithm will assign the role of "ratio denominator" to some predictors (negative-signed b coefficients) and that of "ratio numerator" to others (positive-signed b coefficients) so that $b_1 + b_2 + \cdots = 0$, the overall effect being the removal of size. Items where negative numbers occur will equally form linear combinations capable of removing size as shown in (4). Models ensuing from such general case are capable of portraying financial features required by the relationship at hand. But it is doubtful whether analysts would be able to interpret and use such models.

The conclusion to be drawn at this stage is that the task of selecting ratios capable of portraying financial features is not a modelling pre-requisite. In response to size-independence, the algorithm may form high-dimensional ratios from N logarithmic-transformed items. Only in the presence of exactly $N = 2$ predictors, say, items m and n, will the algorithm form the usual type of ratio, capable of being interpreted and used by analysts. But, as shown further down in the paper, the algorithm can be compelled into forming $N-1$ pairs (differences $\log x_m - \log x_n$), which are logarithmic ratios. Pairing will not bring any new variability into the model, thus no improvement in performance should be expected just by subtracting two logarithmic-transformed line items and then using such subtractions as predictors. The inverse, however, is quite possible: model performance may be hampered when pairing leads to curtailed explanatory variability.

2.3 Size-related prediction

Until this point it was presupposed that y is size-independent. Where y is size-related, models might simply use, as in (6), logarithmic-transformed line items as predictors. The term $(b_1+b_2+\cdots)s$ in (7) apportions the required size-related variability. In this way, the problem of having to select predictors from an excessive number of candidates also ceases to exist as, instead of having some 1,200 ratios to choose from, predictors would be selected from among the comprehensive set of usable line items, some 50 predictors. But such models' coefficients and their signs would be, in analysts' eyes, utterly non-intuitive. In addition, changes in any magnitude-dependent effect such as an altered currency, would require the building of a new model.

A difficulty of a different kind, also associated with (6), is dependence among the $b_1, b_2, \cdots$ created by the common effect s. Size-independent models, where the constraining mechanism at work is $b_1 + b_2 + \cdots = 0$, may account for such mechanism by using $N-1$, not N coefficients, as illustrated above for $N = 2$ and $N = 3$. But size-related models will have their coefficients distorted by $(b_1 + b_2 + \cdots)s \neq 0$. Size, which is present in all line items, will not be wholly offset, thus creating dependencies and exposing algorithms to multicollinearity.

Fortunately, in consideration to analysts' interpretability demands and for model portability reasons, ratios are the sole predictors used in most application tasks, size being thus prevented from entering the model. In the present formulation, the same is accomplished by using differences $\log x_m - \log x_n$. Given the diversity of firms' sizes typical of most cross-sections, the predictive role of size is, in general, small, especially in classification tasks. Little is missed, therefore, by preventing size from entering models.

Still, there are a few application tasks where the effect of size cannot be omitted without significantly affecting model performance, model meaning or both. In such cases, instead of using (6), size can be separated from size-independent effects, as in research-oriented formulations, thus allowing interpretability of coefficients. This is accomplished by using a size proxy together with ratios (pairs $\log x_m - \log x_n$ in the present formulation).

To sum up, given a subset of N line items $x_1, x_2, \cdots$ possessing predictive power over an attribute y, the formal basis just developed suggests the use of modelling formulations

$$y = a + \sum_{k=1}^{N-1} b_k (\log x_m - \log x_n)_k \tag{9}$$

for size-independent y and

$$y = a + \sum_{k=1}^{N-1} b_k (\log x_m - \log x_n)_k + S \tag{10}$$

for size-related y. Pair k is formed from line items m and n. Appropriate $N-1$ pairs should be formed from all the N line items. S is a size proxy, a is the constant term, and log may refer to (5) in the case of positive- and negative-signed x. Notwithstanding the fact that the b_k in (10) refer to size-independent variables, some degree of dependence among coefficients is unavoidable.

2.4 Other modelling requirements

It may be necessary, prior to logarithmic transformation, to scale numbers back into original values, as most databases of financial statements scale numbers to millions, which may make them fall within the range $-1 < x < 1$. Scaling may also be necessary when applying logarithmic transformations to Dividends-per-Share and other pre-existing ratios where individual components are unknown. Scale must be taken into account when applying models but performance and significance statistics will not change with scale.

Besides ratios, relative changes of line items in relation to the previous period are also employed in the analysis of financial statements and as predictors in models. Relative changes should be transformed into subtractions of logarithms,

$$\delta \log x = \log x_t - \log x_{t-1} \tag{11}$$

where t and $t-1$ express subsequent time periods and the operator log may refer to (5) in the case of positive- and negative-signed x. Contrary to relative changes, subtractions (11) will not generate missing values when previous year's value is zero.

3 The proposed methodology

By using $N-1$ pairs $\log x_m - \log x_n$ as an alternative to ratios x_m/x_n, performance will not be endangered by influential cases or by heteroscedasticity, models will be interpretable, and size will not enter the relationship except when explicitly accounted for. But the selection of such pairs raises the same difficulties as the selection of ratios. The proposed methodology addresses such dilemma by performing the modelling process in two stages. At the first stage, the variability required by the modelling of y is identified while still in the form of line items, not pairs. Then, at a second stage, the hidden nodes of an MLP will find adequate pairs from the (much smaller) set of line items selected at the first stage.

3.1 First stage: selecting an optimal set of line items

Variability potentially suitable to model accounting-based relationships is contained in the comprehensive set of line items. Such set may comprise some 50 variables to choose from, being feasible to identify the N line items with predictive power. This may be accomplished by using formulation (6), together with some type of variable-selection capability to be found as part of most modelling algorithms; but dependence among coefficients in size-related models suggests the use of algorithms which are robust regarding correlation among input variables. Here, the algorithm employed is the MLP with one hidden layer, to which input pruning (Cibas et al., 1994) is applied.[2]

Conservatism in variable-selection is essential to the success of the proposed methodology. Typically, least significant variables neither increase nor decrease out-of-sample performance and should be discarded before proceeding to the second stage.

Besides selecting N items from the comprehensive set, the predictive performance level achieved by such N items is an upper-limit to the second stage and should be made a note of.

3.2 Second stage: discovering ratios

It was shown in (8) that, when two logarithmic-transformed line items are set to predict a size-independent relationship, the ensuing model coefficients tend to be symmetrical (opposite signs and approximately similar absolute values). Such two predictors are effectively forming a ratio in logarithmic space. The finding of a set of financial ratios is accomplished by compelling a conveniently-trained MLP to assume an architecture where every node in the hidden layer has no more than two connections to inputs (logarithmic-transformed line items).

First, the N optimal line items identified at the first stage are made present to the MLP together with the corresponding instances of the attribute to be predicted. The training begins with just one node in the hidden layer and the least significant input connections are identified (Garson, 1991), pruned, and the MLP is trained anew. This is repeated until only two input connections remain in the node. It is observed that such surviving connections have opposite signs and broadly similar magnitudes.

The hidden layer is then made to grow to a maximum of, typically, $N-1$ nodes. Every time a new node is included, the two most significant, opposite-signed input connections are identified and the remaining connections are pruned in the way just described.

Where, in a node just entered, the most significant input variable is a pre-existing ratio (Earnings per share, Dividends per Share, the change in relation to the previous period) or an industry dummy, the corresponding connection is preserved and all the others are pruned. The maximum number

[2] Other methods are also available (Stahlberger and Riedmiller, 1996).

of nodes is, in this case, increased by 1.

Where, at the end of the process, performance is markedly inferior to first-stage performance, it is assumed that size cannot be omitted. In such case, the process is repeated, the first node to enter being connected solely to an appropriate size proxy and the maximum number of nodes now being N, not $N-1$.

When growth is completed, the less significant node (Mozer and Smolensky, 1988), which is not necessarily the last node to enter, should be tested for pruning, the *ad hoc* criterion being a non-significant reduction in classification accuracy.

It is easy to tell which ratios have been formed: the signs of connections linking inputs to hidden nodes show which input is acting as the numerator (positive sign) and denominator (negative sign) of the newly discovered ratios. Connections linking hidden nodes to output nodes are interpreted in the way conventional model coefficients are: their magnitudes broadly suggest the relative importance of each node while their signs show in which direction the newly discovered ratios influence output. When size is explicitly included in the relationship, the remaining variables perform size-independent prediction. For the sake of conformity with analysts' usage, in some cases ratios must be inverted, their signs changing accordingly. Indeed, it would be uncommon to present ratios where, for instance, Assets (total) or Liabilities (total) are in the numerator of the ratio.

In what sense is the discovered set of ratios optimal? Since the first-stage model performance is the upper limit on performance, the discovered set of pairs will be optimal if it performs as well as the set of items it is drawn from. It may be argued that the optimal set of pairs may require the use of items which were not selected at the first stage. But, given that the subtraction of two variables uncorrelated to a third variable is uncorrelated to such third variable, it is impossible for two variables to have no predictive power over a given attribute and yet their difference have (the reverse, though, is quite possible).

Most of the available MLP algorithms can be utilised to implement the procedure just described,[3] simply by repeatedly training the MLP while importing, exporting and pruning connections in the described way. Other procedures also leading to good results (for instance, the introduction of competition for survival among connections of an entering node) would require dedicated programming.

[3]For instance, the Stuttgart Neural Network Simulator (SNNS) and associated tools, which can be accessed from within R and other languages.

4 Three application examples

The proposed methodology is now set to predict the following attributes:

1. Bankruptcy, with two states: bankrupt (positive) and non-bankrupt (solvent, negative) (Balcaen and Ooghe, 2006);

2. Financial misstatement, with two states: misstated (manipulated, fraudulent, positive) vs fair (negative) (Ngai et al., 2011);

3. Earnings' increase one year ahead, with two states: increase (positive) vs non-increase (negative) (Bird et al., 2001).

Examples illustrate three levels of difficulty (low, medium and high), having been researched, documented and replicated under varied conditions. Examples include balanced and unbalanced states, matched and unmatched sampling, size-independent and size-related relationships, and both *ex post* and *ex ante* out-of-sample performance assessment.

So far, bankruptcy is the only clearly predictable accounting-based attribute, with reported 95% or higher classification accuracy.[4] Misstatement detection accuracy is at 75% for diversified samples. Besides being meagre, such accuracy is unbalanced, Type I and Type II errors diverging widely and unpredictably. Earnings prediction is an even more elusive task, with barely 60% of correct classification and highly unbalanced accuracy.

Input and target attributes employed for learning and out-of-sample accuracy assessment are taken from the following data-sources:

1. A list of U. S. bankruptcy filings,[5] covering the period 1978-2008. Both Chapter 7 and Chapter 11[6] bankruptcies are included. All bankruptcies but the first in a company are excluded.

2. The list of accounting and auditing enforcement releases, resulting from investigations made by the U. S. Securities and Exchange Commission for alleged accounting misconduct, covering the period 1983-2008.[7]

[4]Classification accuracy figures mentioned here presuppose that states in model-building data are balanced. The use of ensemble modelling typically increases accuracy by 3-5%.

[5]http://lopucki.law.ucla.edu/ (UCLA-LoPucki Bankruptcy Research Database) and other sources

[6]Under Chapter 11 debts are restructured so that debt repayment is possible; under Chapter 7 assets are liquidated to repay debts.

[7]http://groups.haas.berkeley.edu/accounting/faculty/aaerdataset/ (Centre for Financial Reporting and Management, Haas School of Business, University of California at Berkeley)

3. The "Compustat" repository of financial data by Standard & Poor's, from which all monetary amounts are drawn and Earnings' changes are estimated (Ou, 1990).

Such data-sources are easily available, the latter being routinely used by analysts.

The comprehensive set from which, at the end of the first stage, a reduced subset is drawn, is the same for the three examples and consists of 69 variables:

- 29 logarithmic-transformed line items listed in Table 1. A sharp increase in missing values advises against trying to reach a higher level of item disaggregation (Chen et al., 2015);
- the corresponding 29 differences in relation to the previous period, as in (11);
- one pre-existing ratio, Dividends per Share, also included in Table 1, and the corresponding difference in relation to the previous period.
- 9 dummies, one for each of the Global Industry Classification Standard (GICS) sectors.[8] Typically, sector dummies end-up not being selected.

Besides listing variables hypothesised as comprising the comprehensive set, Table 1 also lists the transformations applied to each of them and, for the three examples, the variables selected at the first stage and the roles they assume at the second stage.

The common characteristics of the MLP used at the first and second stages for the three examples, are:

- hyperbolic tangents (threshold functions ranging from -1 to $+1$) are used as transfer functions in all nodes.
- two output nodes; instances to be predicted are symmetrical, $+1$ denoting a positive state (bankruptcy, misstatement or Earnings' increase) and -1 denoting a negative state (non-bankruptcy, non-misstatement or Earnings' non-increase). Signs and magnitudes of connections to output nodes refer to the corresponding states.
- connections are initially set to random values between -1 and $+1$.

[8]https://www.msci.com/gics

Table 1: The 29 line items plus 1 pre-defined ratio from which MLP inputs are selected for the prediction of bankruptcy (BK), misstatement detection (MSA) or the forecast of Earnings' changes (EAF). The transformation applied to each item (TRF) is either logarithmic (log) or logmodulus (lm). Selected variables may be numerators (N) or denominators (D) of ratios, a size proxy (S), changes in relation to previous year (C) or a predefined ratio (R).

Variable	TRF	BK	MSA	EAF
Cash and Short Term Investments	log		N	
Receivables (total)	log		N	
Inventories (net)	log			
Current Assets (total)	log			
Property Plant and Equipment (net)	log			
Investment and Advancements	log			
Intangibles (total)	log			
Assets (total)	log	D	D	D, S
Account Payable (trade)	log			
Accrued Expenses	log			
Current Liabilities (total)	log			
Long-Term Debt	log		N	
Liabilities (total)	log	D, D	N, C	D
Common Stock (equity)	log		D, D	
Preferred Stock	log			
Retained Earnings	lm	N		N, N, C
Shareholders' Equity (total)	lm			
Revenue (total)	log		N, D	
Cost of Goods Sold	log			
Gross Profit	lm			N, C
Operating Expense (total)	log			
Selling, General and Administrative Expenses	log			
Operating Income after Depreciation	lm			
Interest and Related Expense	log			
Income Tax (total)	lm			N, D, D
Net Income (loss)	lm			
Cash-Flow from Operating Activities (net)	lm	N, N		N, D
Cash-Flow from Financing Activities (net)	lm			
Cash-Flow from Investing Activities (net)	lm			
Dividends per Share ex-date (fiscal)	lm			R, C

All reported classification accuracy is out-of-sample, referring, not to MLP performance, but to Logistic Regression performance where logarithmic-transformed variables (first stage) or discovered pairs (second stage) are the predictors. In fact, the aim here is not to report algorithmic performance but performance of the newly-discovered ratios under fairly usual conditions. It would be pointless to use MLP performance to illustrate the benefits of the newly-discovered ratios because such performance depends heavily on architecture, training algorithms and other factors; and also because reported results must be comparable to previously published predictive modelling results where Logistic Regressions are routinely used. Accuracy figures refer to formulation (6) for variable selection (first stage), and to (9) or (10) for size-independent or size-related prediction respectively.

The remaining of the section details, for the three examples, relevant literature and results, sampling details, ratios formed at the second stage and the observed increase in classification accuracy. A *ceteris-paribus* interpretation of connections linking hidden nodes to output nodes is also offered.

4.1 Bankruptcy prediction

The first example, bankruptcy prediction, is inspired by an earlier, popular attempt to model a financial attribute (Altman, 1968). An extensive review of bankruptcy-predicting models is available (Balcaen and Ooghe, 2006).

After discarding filings for which financial statements are not available, two random samples of nearly 900 filings each are drawn. All sizes (deciles of the logarithm of Assets), years, and selected GICS sectors are represented in such samples.

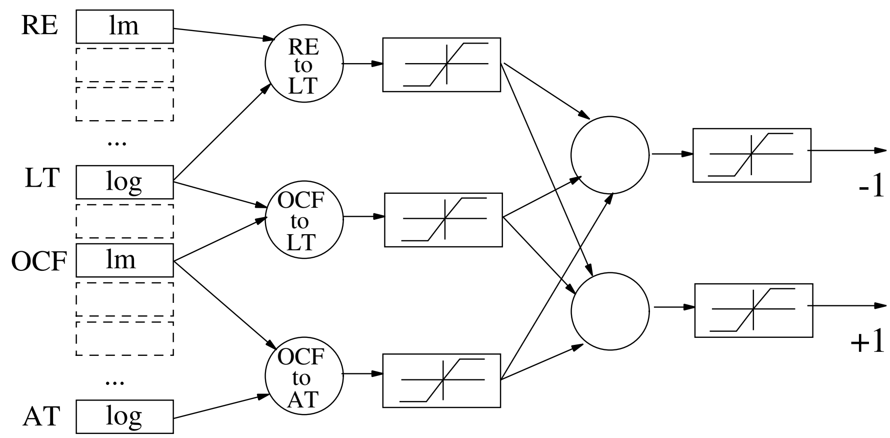
Each case in the two samples is matched with a non-bankrupt case. Matching is based on the GICS sector, on size decile and on year. One bankrupt case is randomly assigned to its peer and then the peer is made unavailable for future matching. One of the two datasets thus obtained is utilised to build (learn, train) models and the other to test model's accuracy. Due to missing data, the size of such datasets is less than 1,800:

Learning-set: non-bankrupt	786 (50%)
Learning-set: bankrupt	787 (50%)
Test-set: non-bankrupt (N)	796 (50%)
Test-set: bankrupt (P)	795 (50%)

Only 4 input variables are selected at the first stage (Table 1).

During the second stage, 3 ratios are formed in the MLP's hidden layer:

Figure 1: Bankruptcy-prediction second-stage MLP. RE: Retained Earnings, LT: Liabilities (total), OCF: Cash-Flow from Operating Activities, AT: Assets (total), log: logarithmic transformation, lm: logmodulus transformation.



Node 1 – Cash-Flow from Operating Activities (net) to Liabilities (total) - negative output connection: other things being equal, the higher this ratio is, the less likely bankruptcy is. Accuracy gain is 37.4%.

Node 2 – Retained Earnings to Liabilities (total) - negative output connection: other thins being equal, the higher this ratio is, the less likely bankruptcy is. The gain in accuracy brought about by this node is 5.8%.

Node 3 – Cash-Flow from Operating Activities (net) to Assets (total) - positive output connection: other things being equal, the higher this ratio is, the more likely bankruptcy is. The gain in accuracy brought about by this node is 3.7%.

In each of the three hidden nodes, input connections are of crudely similar magnitudes and opposite signs as predicted by theory. Figure 1 shows the MLP architecture.

When the 4 selected variables (first stage) or the 3 newly found pairs (second stage) are used as predictors in Logistic Regressions, classification accuracy (Table 2) is similar to MLP accuracy. Performance is not inferior to that reported in the literature for similar conditions, while the number of

Table 2: Example 1: bankruptcy prediction. The table shows sample sizes and out-of-sample accuracy. T: number of true (correctly classified) cases, F: false (incorrectly classified), P: positives (bankrupt), N: negatives (non-bankrupt).

		first-stage (4 var.)		second-stage (3 var.)	
Accuracy	$T/(T+F)$	1,547	(97.2%)	1,541	(96.9%)
No bankrupt correct	TN	780	(98.0%)	778	(97.7%)
No bankrupt incorrect	FP	16	(2.0%)	18	(2.3%)
Bankrupt correct	TP	767	(96.5%)	763	(96.0%)
Bankrupt incorrect	FN	28	(3.5%)	32	(4.0%)
Precision	$TP/(TP+FP)$	98.0%		97.7%	
Sensitivity	$TP/(TP+FN)$	96.5%		96.0%	

predictors is smaller. The observed reduction in performance from the first to the second stage is negligible, evidencing the non-significance of size.

4.2 Financial misstatement detection

The second example replicates misstatement-detection models. Research published in Accounting and Finance journals (Dechow et al., 2011) and in Artificial Intelligence journals (Huang et al., 2014) report accuracy ranging from 65% to 75% for large, non-homogeneous samples whereas, for small, same-industry samples, accuracy can be as high as 86% (Huang et al., 2014). In all such cases Type I and Type II errors may differ by as much as 30%.

The dataset is divided in two sets, one containing misstatement releases issued from years 1976 to 1999 and the second containing releases issued from years 2000 to 2008. The 1976-1999 set is utilised for model building and MLP learning; the 2000-2008 set is utilised to measure accuracy. The two sets are matched with an equal number of firms which are neither the object of releases throughout the period nor are bankrupt in the same year. Matching is based on the GICS sector, on size decile and on year. Each misstatement case is randomly assigned to its peer and then the peer is made unavailable for future pairing. The resulting datasets have nearly 900 cases each but, due to the existence of missing values, the number of useful cases is smaller:

Learning set: non-fraud cases	335	(45.7%)
Learning set: fraud cases	398	(54.2%)

Test set: non-fraud cases (N)	353 (46.2%)
Test set: fraud cases (P)	411 (53.8%)

A total of 8 variables are selected at the first stage, of which 7 are transformed line items and one is a difference in relation to the previous period (Table 1). It is verified that the relationship is size-independent.

During the second stage, 6 hidden nodes are formed:

Node 1 – Receivables (total) to Common Stock (equity) - negative connection to output node: other things being equal, the higher this ratio is, the less likely misstatement is. Accuracy gain is 27.2%.

Node 2 – Liabilities (total) to Assets (total) - positive connection to output node: other things being equal, the higher this ratio is, the more likely misstatement is. Accuracy gain brought about by this node is 4.8%.

Node 3 – Cash and Short Term Investments to Revenue (total) - negative connection to output node: other things being equal, the higher this ratio is, the less likely misstatement is. Accuracy gain brought about by this node is 3.4%.

Node 4 – Long Term Debt to Common Stock (equity) - negative connection to output node: other things being equal, the higher this ratio is, the less likely misstatement is. Accuracy gain brought about by this node is 1.2%.

Node 5 – Change in Liabilities (total) - positive connection to output node: other things being equal, an increase in liabilities relative to previous period demotes higher likelihood of misstatement. Accuracy gain brought about by this node is 0.8%.

Node 6 – Revenue (total) to Common Stock (equity) - negative connection to output node: other things being equal, the higher this ratio is, the less likely misstatement is. This node actually reduces accuracy by 0.6%.

The fifth node added to the MLP has only one input connection. Other nodes have two input connections with broadly similar magnitudes and opposite signs as predicted.

When the 8 selected variables (first stage) or the 6 newly found pairs (second stage) are used as predictors in Logistic Regressions, classification accuracy (Table 3) is similar to that of MLP accuracy. Notably, in this

Table 3: Example 2: misstatement detection. The table shows sample sizes and out-of-sample accuracy. T: number of true (correctly classified) cases, F: false (incorrectly classified), P: positives (misstated), N: negatives (not misstated).

		first-stage (8 var.)		second-stage (6 var)	
Accuracy	$T/(T+F)$	764	(88.2%)	668	(87.4%)
No fraud correct	TN	303	(85.8%)	299	(84.7%)
No fraud incorrect	FP	50	(14.2%)	54	(15.3%)
Fraud correct	TP	371	(90.5%)	369	(90.0%)
Fraud incorrect	FN	39	(9.5%))	41	(10.0%)
Precision	$TP/(TP+FP)$	88.1%		87.2%	
Sensitivity	$TP/(TP+FN)$	90.3%		89.8%	

case accuracy is around 10% higher than that reported in the literature for similar conditions. Imbalance in the recognition of states is smaller while Type II error (the most expensive in this case) is clearly subdued. Reduction in performance from the first to the second stage is negligible, denoting size-independence.

4.3 Earnings forecasting

The third example is about the forecasting of Earnings' changes one year ahead. Its distinctive marks are the use of unmatched, unbalanced datasets, the large number of available cases (states can be computed from specific line items), and a weak relationship, with a reported accuracy of barely 10% above a prediction made at random.

After computing the two stages to be predicted (Ou, 1990), cases with missing values are put aside. A total of nearly 140,000 cases remain, where some 90,000 are Earnings' non-increases and 50,000 are increases. From this unbalanced set, two samples are drawn, one for learning and the other for performance assessment:

Learning set: Earnings' non-increases	41,851	(64.3%)
Learning set: Earnings' increases	23,275	(35.7%)
Test set: Earnings' non-increases (N)	41,750	(64.4%)
Test set: Earnings' increases (P)	22,811	(35.6%)

Net Income (loss) is omitted from the comprehensive set in this case (Ou, 1990). A total of 10 variables are selected at the first stage, 3 of which are differences in relation to the previous period, 1 is a predefined ratio (Dividends per Share), and the remaining 6 are logarithmic-transformed line-items.

It is further verified that the relationship is significantly size-related. Therefore, the first node to enter is connected solely to the logarithm of Assets (total), a suitable size proxy which is selected at the first stage. The number of hidden nodes is not $N-1=9$ but 10. Nodes enter in the following order:

Node 1 – Assets (total) - negative connection to output node: other things being equal, larger companies are less likely to report an increase in Earnings. This node brings about an accuracy gain of 5.3% above the class imbalance of 14.7%.

Node 2 – Dividends per Share - negative connection to output node: other things being equal, the more dividends per share a company pays, the less likely an Earnings' increase is. This node brings about an accuracy gain of 7.1%

Node 3 – Cash-Flow from Operating Activities (net) to Income Tax (total) - negative connection to output node: other things being equal, the higher this ratio is, the less likely an Earnings' increase is. This node brings about an accuracy gain of 2.5%

Node 4 – Retained Earnings to Liabilities (total) - negative connection to output node: other things being equal, the higher this ratio is, the less likely an Earnings' increase is. This node brings about an accuracy gain of 0.8%

Node 5 – Change in Gross Profit - positive connection to output node: other things being equal, an increase in Gross Profits relative to previous period denotes higher likelihood of an Earnings' increase.

Node 6 – Retained Earnings to Income Tax (total) - positive connection to output node: other things being equal, the higher this ratio is, the more likely an Earnings' increase is.

Node 7 – Gross Profit to Cash Flow from Operating Activities (net) - negative connection to output node: other things being equal, the higher this ratio is, the less likely an Earnings' increase is.

Node 8 – Income Tax to Assets (total) - positive connection to output node: other things being equal, the higher this ratio is, the more likely an increase in Earnings is.

Table 4: Example 3: Earnings increase forecasting. The table shows sample sizes and out-of-sample accuracy. T: number of true (correctly classified) cases, F: false (incorrectly classified), P: positives (Earnings' increases) N: negatives (Earnings' non-increases).

		first-stage (10 var.)		second-stage (10 var.)	
Accuracy	$T/(T+F)$	51,936	(80.4%)	51,587	(79.9%)
No incr. correct	TN	35,783	(85.7%)	35,592	(85.3%)
No incr. incorrect	FP	5,967	(14.3%)	6,158	(27.0%)
Increase correct	TP	16,153	(70.8%)	15,995	(70.1%)
Increase incorrect	FN	6,658	(29.2%)	6,816	(16.3%)
Precision	$TP/(TP+FP)$	73.0%		72.2%	
Sensitivity	$TP/(TP+FN)$	70.1%		70.1%	

Node 9 – Change in Retained Earnings - positive connection to output node: other things being equal, an increase in Retained Earnings relative to previous period denotes a higher likelihood of an Earnings' increase.

Node 10 – Change in Dividends per Share - positive connection to output node: other things being equal, an increase in dividends per share relative to previous period denotes higher likelihood of an Earnings' increase.

Marginal classification accuracy ceases to be meaningful from node 5 onwards, overall accuracy neither increasing nor decreasing. Accuracy becomes unreliable as a guiding signal, except for the most significant pairs. But when the 10 selected variables (first stage) or the 10 predictors (second stage) are used in Logistic Regressions, all such variables are statistically significant ($P < 0.01$). Overall classification accuracy (Table 4) is, in this case too, broadly similar to MLP accuracy.

Classification results should be interpreted in the light of state imbalances observed in the learning-set (Chawla, 2005). Performance increases by some 6% in average relation to previously reported research. Importantly, in the presented example state imbalance is similar to that in the dataset employed to learn the relationship.

4.4 Discussion

It would be difficult to contradict the assertion that, for fairly typical conditions, the set of variables selected from the comprehensive set at the first stage is an optimal predicting set. And, provided that due parsimony is observed in such selection, pairs formed in the hidden node of the second-stage MLP will indeed exhibit opposite signs and broadly similar magnitudes. As a consequence, the set of ratios discovered at the second stage, exhibiting a predicting power similar to that of the first stage, will indeed be an optimal set as well.

The three examples just provided are linear relationships with non-significant interactions between predictors. This is the usual scenario for the type of applied models analysts use. The proposed methodology is not meant to discover ratios under conditions or in application areas where financial analysis has little concern. Therefore, besides not covering higher-order relationships, the proposed methodology is also not meant to discover ratios for rare events modelling (King and Zeng, 2001) or other extreme tasks. Namely, states to be predicted should be, to a reasonable extent, likely states with easily available patterns allowing their recognition.

It may be argued that the matching of cases, as in examples 1 and 2, or other undefined causes, may have led to a reduction in size-dependence in the attribute to be predicted. In turn, such reduction would facilitate the selection of variables at the first stage and the forming of pairs at the second stage. It may also be speculated that a given set of predictors may be optimal for a given state imbalance, say 80%-20%, and non-optimal for another; or even that different state-imbalances may require different predictors. These are interesting questions for theory-driven research. When the modelling algorithm is the Logistic Regression, it was demonstrated that state-imbalance affects the model intercept only (Yu and Manski, 1989), and an intercept correction is available (King and Zeng, 2001), making it possible to build a model using balanced states (for example, by matching) and then correct the intercept to take real-world imbalance into account. Moreover, in example 3 appropriate variables were selected and pairs formed, in spite of matching not being used. Given this, it seems as though the discovery of sets of predictors and their optimality should be viewed as robust, within a reasonable range, regarding differences in state imbalance.

5 Conclusion

If ratios can be validly employed in financial analytic tasks, then the effect of company size, which ratios remove, can also be removed by modelling algorithms. While removing size, algorithms portray financial features required by the relationship being modelled, as though ratios were there.

The paper suggested that predictors to be used in accounting-based models should be selected by the algorithm from among transformed line items, not from ratios possessing suitable analytical qualities. Models should be built in two stages, one to find the set of transformed items able to explain the relationship optimally, and the other to form, from such set, appropriate pairs of predictors. Such pairs are, in fact, ratios in logarithmic space. A procedure was suggested to automatically discover an optimal set of pairs. The reported examples suggest that the proposed procedure is feasible, bearing fruit when applied to modelling problems where the finding of appropriate ratios is non-intuitive, as is the case of misstatement-detection or earnings' changes forecasting.

The paper illustrates a case of close alignment between statistical adequacy, users' needs and algorithmic architecture. The choice of the knowledge-extraction algorithm, the MLP, was dictated solely by its ability to form internal representations. Neither performance nor the testing of novel capabilities was the goal here.

Performance reported in the paper is high for two reasons. First, the logarithmic transformation agrees with the multiplicative character of predictors; appropriately transformed variables, not the algorithm, led to the discovery of logarithmic ratios and then to a parsimonious and balanced prediction. Second, ratios are discovered by the optimisation algorithm. In this way, unduly circumscribed ratios are avoided, meaningful as they may seem to be.

Acknowledgements

This research is sponsored by the Foundation for the Development of Science and Technology (FDCT) of Macau, China. Sponsorship No. 044-2014-A1.

References

Altman, E. (1968) 'Financial ratios, discriminant analysis and the prediction of corporate bankruptcy', *The Journal of Finance*, Vol. 23, No. 4, pp.589–

609.

Altman, E. and Sabato, G. (2007) 'Modelling credit risk for smes: evidence from the u.s. market', *Abacus*, Vol. 43, No. 3, pp.332–357.

Amani, F. and Fadlalla, A. (2017) 'Data mining applications in accounting: a review of the literature and organizing framework', *International Journal of Accounting Information Systems*, Vol. 24, pp.32–58.

Angelini, P. and Generale, A. (2008) 'On the evolution of firm size distributions', *The American Economic Review*, Vol. 98, No. 1 pp.426–438.

Balcaen, S. and 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*, Vol. 38, No. 1, pp.63–93.

Berry, R. and Trigueiros, D. (1993) 'Applying neural networks to the extraction of knowledge from accounting reports: a classification study', *in* R. Trippi and E. Turban (eds), *Neural Networks in Finance and Investing*, Probus Publishing (Chicago, Ill), chapter 6, pp.103–123.

Bird, R., Gerlach, R. and Hall, A. (2001) 'The prediction of earnings movements using accounting data: an update and extension of Ou and Penman', *Journal of Asset Management*, Vol. 2, No. 2, pp.180–195.

Cabral, L. and Mata, J. (2003) 'On the evolution of the firm size distribution: facts and theory', *The American Economic Review*, Vol. 93, No. 4, pp.1075–1090.

Cao, Q. and Parry, M. (2009) 'Neural network earnings per share forecasting models: a comparison of backward propagation and the genetic algorithm', *Decision Support Systems*, Vol. 47, No. 1, pp.32–41.

Chawla, N. (2005) 'Data mining for imbalanced datasets: an overview', *in* O. Maimon and L. Rokach (eds), *Data Mining and Knowledge Discovery Handbook*, (Springer, New York NY), pp.853–867.

Chen, S., Miao, B. and Shevlin, T. (2015) 'A new measure of disclosure quality: the level of disaggregation of accounting data in annual reports', *Journal of Accounting Research*, Vol. 53, No. 5.

Cibas, T., Fogelman, F., Gallinari, P. and Raudys, S. (1994) 'Variable selection with optimal cell damage', *International conference on Artificial Neural Networks, ICANN*, pp.727–730.

Cinca, C. (1998) 'From financial information to strategic groups – a self-organizing neural network approach', *Journal of Forecasting* Vol. 17, pp. 415–428.

Dechow, P., Weill, G., Larson, C. and Sloan, R. (2011) 'Predicting material accounting misstatements', *Contemporary Accounting Research*, Vol. 28, No. 1, pp.17–82.

Garson, G. (1991) 'Interpreting neural network connection weights', *Artificial Intelligence Expert*, Vol. 6, No. 4, pp.46–51.

Hong, T. and Shin, T. (2007) 'Developing corporate credit rating models using business failure probability map and analytic hierarchy process', *The Journal of Information Systems*, Vol. 16, No. 3, pp.1–20.

Huang, S., Tsaih, R. and Yu, F. (2014) 'Topological pattern discovery and feature extraction for fraudulent financial reporting', *Expert Systems with Applications*, Vol. 41, No. 9, pp.4360–4372.

Ijiri, Y. and Simon, H. (1964) 'Business firm growth and size', *The American Economic Review*, Vol. 54, No. 2, pp.77–89.

Ijiri, Y. and Simon, H. (1974) 'Interpretations of departures from the pareto curve firm-size distributions', *Journal of Political Economy*, Vol. 82, No. 2, pp.315–331.

John, J. and Draper, N. (1980) 'An alternative family of transformations', *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, Vol. 29, No. 2, pp.190–197.

King, G. and Zeng, L. (2001) 'Logistic regression in rare events data', *Political Analysis*, Vol. 9, No. 2, pp.137–163.

Landajo, M., de Andrés, J. and Lorca, P. (2007) 'Robust neural modeling for the cross-sectional analysis of accounting information', *European Journal of Operational Research*, Vol. 177, No. 2, pp.1232–1252.

McLeay, S. and Trigueiros, D. (2002) 'Proportionate growth and the theoretical foundations of financial ratios', *Abacus*, Vol. XXXVIII, No. 3, pp.297–316.

McNelis, P. (2004) *Neural Networks in Finance: Gaining Predictive Edge in the Market*, Advance Finance, Academic Press.

Mozer, M. and Smolensky, P. (1988) 'Skeletonization: a technique for trimming the fat from a network via relevance assessment', *Neural Information Processing Systems*, Vol. 1, Morgan Kaufmann.

Ngai, E., Hu, Y., Wong, Y., Chen, Y. and Sun, X. (2011) 'The application of data mining techniques in financial fraud detection: a classification framework and an academic review of literature', *Decision Support Systems*, Vol. 50, No. 3, pp.559–569.

Nikkinen, J. and Sahlström, P. (2004) 'Distributional properties and transformation of financial ratios: the impact of the accounting environment', *Advances in International Accounting*, Vol. 17, pp.85–101.

Ou, J. (1990) 'The information content of non-earnings accounting numbers as earnings predictors', *Journal of Accounting Research*, Vol. 28, No. 1, pp.144–163.

Peat, M. and Jones, S. (2012) 'Using neural nets to combine information sets in corporate bankruptcy prediction', *International Journal of Intelligent Systems in Accounting and Finance Management*, Vol. 19, No. 2, pp.90–101.

Rumelhart, D., Hinton, G. and Williams, R. (1986) 'Learning internal representations by error propagation', *Parallel Distributed Processing*, Vol 1, The MIT Press.

Stahlberger, A. and Riedmiller, M. (1996) 'Fast network pruning and feature extraction by using the unit-obs algorithm', *NIPS*, pp. 655–661.

Sustersic, M., Mramor, D. and Zupan, J. (2009) 'Consumer credit scoring models with limited data', *Expert Systems with Applications*, Vol. 36, No. 3, pp.4736–4744.

Trippi, R. and Turban, E. (eds) (1992) *Neural Networks in Finance and Investing: Using Artificial Intelligence to Improve Real World Performance*, McGraw-Hill (New York NY).

Yu, X. and Manski, C. (1989) 'The logit model and response-based samples', *Sociological Methods and Research*, Vol. 17, No. 3, pp.283–302.