The use of Financial Ratios for Research: Problems Associated with and Recommendations for Using Large Databases

by

Boris Nenide
School of Business
University of Wisconsin-Madison
(608) 265-5080

Robert W. Pricer
Graham Wisconsin Distinguished Professor
School of Business
University of Wisconsin-Madison
(608) 263-3464

S. Michael Camp
Vice-President of Research
Kauffman Center for Entrepreneurial Leadership
Kauffman Foundation
(816) 932-1168

The use of Financial Ratios for Research:
Problems Associated with and Recommendations for Using Large Databases

Abstract

A review of published research in the field of accounting and finance reveals that the use of ratio calculations with multivariate analysis for predicting the performance of business firms is common. However, much of this research uses large database information without determining if needed sample assumptions can be met for reliable conclusions to be drawn by the researchers. This paper presents recommended adjustment techniques for researchers using large databases for ratio calculation to allow for confidence that the results of analysis will be meaningful and that inferences may be drawn from the data. Using a sample from the Kauffman Center for Entrepreneurial Leadership Financial Statement Database, Balance Sheet and Income Statement data of 250 firms is used to illustrate and explain techniques for data error identification, handling the problem of denominators being negative or approaching zero when calculating ratios, and effective techniques for transforming the data to achieve approximation of normal distributions. The application of these recommendations will allow researchers to use financial statement data samples that will meet required characteristics for the use of valid multivariate statistical analysis.

Introduction

The use of financial ratio analysis for understanding and predicting the performance of privately owned business firms is gaining in importance in published research. Perhaps the major problem faced by researchers is the difficulty of obtaining an adequate sample of representative financial statements with many studies using 50 or fewer firms for analysis. However, when larger databases are used, it is important to know that they have problems as well and that adjustments to these samples must be made to permit the use of multivariate analysis techniques.

Understanding how to properly use large databases for ratio analysis will become of importance now that the Kauffman Center for Entrepreneurial Leadership (KCEL) has developed a financial statement database of more than 400,000 privately owned firms with a significant number of these including a base year and three operating years of financial statements. This database is currently available to a team of scholars working closely with the KCEL on selected internal studies. It is expected that this database will become generally available to researchers and this source of financial statement information is likely to become the standard for financial performance research in the future. For the first time, scholars will have a large commonly available database of privately owned firm financial statements that will provide direct comparisons between research findings and research formulation.

The advantage of having a large database that is readily available to researchers is easily understood. However, it is equally important to know the shortcomings and needed adjustments of large databases if meaningful research findings are to be achieved. The problems common to large databases are important and are part of almost any sample of financial statements used for financial or ratio analysis.

The need for reliable financial statement data and the importance of financial ratios for analysis and prediction is well established in the literature. Beginning with Beaver's (1966) contention that standard financial ratios can predict the financial performance of firms, many subsequent studies have attempted to demonstrate the predictive value of various techniques for estimating actual business performance.

A review of the literature describing methods and theories for evaluating and predicting financial performance reveals that although methods have become increasingly complex, few researchers adequately address the problems associated with the sample used. For example, most ratio analysis studies use multivariate analysis that is based on the assumption of a normal distribution of the financial ratios. Without confirming the approximation of normality of ratio distribution, the researcher is at risk of drawing erroneous inferences. When considering the distribution of financial ratios in any database, the normality of the distribution can be skewed by data recording errors, negative denominators and denominators approaching zero (Foster, 1986).

It appears that the more that is understood by researchers about financial performance, the greater the number of subjective and objective variables a particular predictive model is likely to include and the less attention is paid to sample weaknesses. It is common today to read about financial performance predictive techniques that are based on such things as "Inductive Learning" (Shaw and Gentry, 1988), " "Knowledge Based Decision Systems"(Stohr, 1986), "Polytomous Probit Analysis" (Dietrich and Kaplan, 1982), "Recursive Learning" (Marais, Patell and Wofson, 1984) and more recently the "Fair Issac Small Business Credit Scoring Model" (Asch, 1995). The list of such financial performance predictive systems is extensive and each of these techniques attempts to explain and systematize the financial performance evaluation process. However, these techniques are generally referred to as "Expert Systems" (Duchessi, Shawky and Seagle, 1988) and they all combine subjective information unique to a given business as well as financial ratio analysis. Unfortunately, none of these new systems are based on error free data samples and, as a consequence, none have been proven to be reliable in terms of identifying future financial performance. In fact, evidence has been reported that indicates that the samples are not normally distributed and include inaccurate data. As such, these methods may not be of use to managers or educators as techniques for better managing businesses (Pricer and Johnson, 1996).

While these new techniques appear complex on the surface, when the models are examined in detail, the variables contained in them are very basic and easy to understand and apply. Also, while many writers agree that objective measures should be used in place of subjective measures for financial performance prediction (Hyndman, 1994), others argue that the only way to assess future financial performance is through the inclusion of subjective measures (Malhorta and McLeod, 1994). However, all of the techniques suffer from sample data that does not meet the requirement of approaching normal distribution or for data accuracy (Kim, 1997; Pricer and Johnson, 1996; Fernandez-Castro, 1994).

When the systems for predicting future financial performance of firms are viewed closely, it is common to find individual variables to be easy to understand and apply. When this is done the apparent weaknesses in the system are detected. For example, one system (Shaw and Gentry, 1988) presents an 80 variable model that includes financial leverage as one measure. Under this model, low risk firms have low financial leverage ratios and high-risk firms have high financial leverage ratios. While this follows common accounting belief, this simplistic use of a variable does nothing to lead to an accurate prediction of firm performance or better management practices. For many of these systems, if the numerator and denominator are negative the ratio calculation is positive. This means that a firm that has negative owner's equity and a loss will have a positive return on equity calculation. This type of error, while common in many studies and systems designed to analyze or predict financial performance, are due to errors in the sample data being used without needed adjustments.

It should be noted that these systems are not accurate in predicting financial performance of firms or for even predicting single ratios for the future (Pricer and Johnson, 1996), but they are increasingly being used by analysts to understand and to forecast firm performance.

The problem encountered is the unfounded belief that raw financial data ratio calculations will lead to valid measures for predicting firm performance. This area of research and theory is generally recognized as having started with liquidity research beginning with Beaver (1966). Beaver tested the ability of 30 standard accounting ratios, four of them cash flow based, to predict the failure of a business as early as five years in advance. These ratios were tested on a sample of 79 failed firms. Beaver concluded that ratios, especially those that measure cash flow coverage of debt, could predict the failure of a business as early as five years in advance. This study, and a number of others that follow, does not discuss whether the sample data meets needed assumptions for normality or if errors were removed from the data before analysis. Altman (1981) attempted to improve conventional ratio analysis by using multivariate analysis on a sample of manufacturing firms, 105 bankrupt firms and 2,058 nonbankrupt firms. Ohlso (1980) concluded from his research that firm size was directly related to firm financial performance with smaller firms more likely to fail than larger ones.

Zavgren (1985), using a sample of 45 bankrupt and 45 nonbankrupt firms, identified seven variables that were used to predict the future financial performance of businesses. Deakin (1972) advanced the research of Beaver and Altman by including the fourteen important ratios identified by Beaver with the multivariate methodology of Altman. Using a sample of 32 failed and 32 nonfailed firms, Deakin found that cash flow coverage to total debt was important for predicting bankruptcy. Blum (1974) also used a failed versus nonfailed model in his research for predicting bankruptcy of a firm.

All of these authors, Beaver (1966), Altman (1981), Ohlso (1980), Zavgren (1985), Deakin (1972) and Blum (1974) can be faulted for using samples that have not been checked for normal distribution or for the removal of errors. In addition, many have inappropriately used mixed or heterogeneous samples in their research. In addition, a close look at the measures causes additional concern. For example, Altman's Z Score uses ratios to predict bankruptcy of a firm but substitutes book value for market value of equity for privately held firms. This variable is given weight in the Z Score calculation and there is no evidence to suggest that book value of equity is in any way equated with the market equity value of a firm (Shah and Murtaza, 2000).

Following the preceding studies, many additional research projects were undertaken in an attempt to validate the use of financial ratios for predicting financial performance of a firm. Some of the better-known studies include Altman, Haldeman and Narayanan (1977), Norton and Smith (1979), and Mensah (1983). These studies, like their predecessors, fail to demonstrate that normality of distribution or that necessary sample assumptions have been met prior to analysis. Even in research that addresses the distribution problem, sample data is transformed without an explanation as to specifically how and why this has been done (Pinches, Mingo and Caruthers, 1973).

During the 1980s, the research emphasis in the area of ratio analysis turned to cash flow indicators following the study of Largay and Stickney (1980) of the failure of W. T. Grant. This largely single case study found that liquidity ratios and measures of cash flows from operations were the best predictors of the future success of a business. However, the conclusions of this study were questioned by the findings of Casey and Bartzca (1984 and 1985). Using a sample of 30 bankrupt firms, with another thirty firms held out for validation, Casey and Bartzca found that standard accounting ratios were better for predicting firm failure than cash flow measures. Unfortunately, the sample assumptions were not tested and this study did not take into consideration firm size when reaching conclusions.

In another study, Gentry (1985) used a sample of 33 bankrupt and 33 nonbankrupt firms to determine if cash flow measures could be used to predict firm financial performance. This study was expanded two years later (Gentry, Newbold, and Whitford, 1987) by testing the ability of both accrual and cash flow measures to predict business financial performance with debatable results. Like many of the other studies cited, the studies also fail to provide evidence that the sample was tested for errors or if it met needed assumptions for approximating a normal distribution. Aziz, Emmanuel and Lawson (1988) combined accrual and cash flow variables in an attempt to predict firm financial performance. However, the results of their validation holdout group and failure to meet needed sample assumptions casts question on their conclusions.

While this literature review is not exhaustive, it does represent the major work that has been done in the field of financial ratio analysis for predicting firm performance. These studies are consistent in their failure to test for sample assumptions and to remove errors for the sample data being used.

The record of the literature is clear; sample data used in ratio analysis must be corrected for errors and tested to determine if normal distribution is approximated. Without these steps, multivariate analysis cannot reliably be performed and results cannot be generalized to a larger population of businesses.

The problem of errors and the inability to meet sample assumptions is present in large databases as well as the smaller samples usually reported in the literature. A recent article provides convincing evidence that data items stored in large databases have significant rates of errors (Klein, Goodhue and Davis, 1997). The problem is exemplified by the COMPUSTAT and CRSP Monthly Return Tape which are mistakenly thought to be accurate because they report information for publicly owned companies. However, both of these databases have significant error rates that distort analysis unless corrected (Kim, 1997; Courtnay and Keller, 1994; Kinny and Swanson, 1993; Bennin, 1980; Beedles and Simkowitz, 1978; and, Rosenberg and Houglet, 1974).

The KCEL financial statement database provides an illustration of how large databases should be modified before ratio analysis. While this is clearly the best database of financial information available, adjustments to it must be made before reliable research conclusion can be reached.

When performing ratio analysis, the following problems need to be accounted for and addressed in the research design if the results are to be meaningful and useful to researchers:

1. Data Entry Errors
2. Negative Denominators
3. Outlier Influence
4. Normality of Distribution

Data Entry Errors:

The financial statement information included in the KCEL financial statement database includes the following Balance Sheet and Income Statement items

Balance Sheet

Assets: Cash, Accounts Receivable, Notes Receivable, Inventory, Other Current Assets, Total Current Assets, Fixed Assets, Other Non-Current Assets, and Total Assets.

Liabilities and Equity: Accounts Payable, Bank Loans, Notes Payable, Other Current Liabilities, Long Term Liabilities, Deferred Credits, Net Worth, and Total Liabilities and Net Worth.

Income Statement
Net Sales, Gross Profit After Tax, Dividends/Withdrawals, and Net Income.

The database also includes the following fourteen standard financial ratios for each firm:

Quick Ratio, Current Ratio, Current Liabilities to Net Worth, Current Liabilities to Inventory, Total Liabilities to Net Worth, Fixed Assets to Net Worth, Collection Period, Sales to Inventory, Assets to Sales, Sales to Net Working Capital, Accounts Payable to Sales, Return on Sales, Return on Assets, and Return on Net Worth.

Based on the data in the financial statement database, we found significant differences between the median and the mean for the fourteen ratios, the standard deviation was very large, and there was no stability of single ratios over time.

For example, when looking at 250 wholesale firm financial statements for the base year plus three years, 1,000 individual statements were available for analysis (base year of 1993 and year-end statements for 1994, 1995 and 1996). The database of these firms includes individually calculated financial ratios and aggregate average ratios for the sample. Many investigators might assume that the information and calculations are accurate when working with such a sample. However, errors were discovered when the statements and calculations were carefully reviewed including 101 statements with zero or negative ratio denominators and 289 statements with errors in ratio calculations.

This illustrates that researchers cannot assume that the financial statement data in a sample is error free. Calculations made without taking the time to check for accuracy will lead to inaccurate findings and conclusions and it is essential that databases be corrected before analysis for this reason.

During the error correction process, it must be remembered that every statement eliminated due to error weakens the representative nature of the sample. However, correction must be made for the data to be useful for detailed analysis and accurate interpretation. While listwise deletion was applied to the sample of this study, which fully removes all records containing obviously erroneous data points, an alternative to this is a smoothing technique called mean substitution which replaces the erroneous data points with either a simple mean or weighted average, either within a particular industry or using other company records. Table 1 presents sample mean, median and standard deviations, both before and after correction. Prior to correction, the distributions are nonsymmetrical, skewed and characterized by very high peaks with extremely flat tails. The corrections generally bring the mean and median much closer together, reduce the standard deviation and some ratios begin to approximate a normal distribution. As Table 1 illustrates, almost all of the ratio calculations are significantly changed when statements with obvious errors are removed from the sample or errors in ratio calculation are corrected.

Table 1 About Here

Negative Denominators

Negative denominator observations will distort not only the analysis of an individual firm's performance, but also the average ratio calculations. For example, if owner's equity is negative, calculating return on equity will have no real meaning. If net income is also negative, the return on equity calculation will be a positive percent and this should not be included in the average calculation figure. Closely related to this problem is the situation where the denominator approaches zero and this may result in high ratios that confound the average ratio calculation. Once again, if owner's equity is positive, but close to zero, a small net income will provide a very high return on equity calculation. Because we assume that comparative ratios have proportional denominators and numerators, deviations from this assumption can lead to unusable ratio calculations.

When faced with the problem of negative or approaching zero denominators, investigators must make a decision as to whether this represents a reasonable part of the financial distribution that should be included or whether it is abnormal and should be deleted. For example, Foster (1986) reports that RMA deletes all negative earnings before taxes data when calculating average profit before taxes to tangible net worth ratios. It must be remembered that each financial statement deletion weakens the representative ness of the sample, but adjustments must be made if the calculations are to be meaningful. This is a difficult decision that can only be made by examining each statement in question and using judgment as to whether it should be included or deleted. Table 2 presents a summary of the sample means, median and standard deviation of the financial ratios before and after error correction and adjustment for negative or approaching zero denominators. This generally brings the mean closer to the median and reduces the standard deviation while providing more stability over time.

Table 2 About Here

Outlier Influence

Before determining if the ratio being studied approximates a normal distribution, the impact of extreme outliers needs to be examined to assess whether they are representative of actual firm performance or if some other circumstance explains the abnormality. A decision must be made to determine if the outlier represents an extreme but consistent indicator of performance. If it does not appear to represent actual firm performance, it must be deleted.

The most common reason for errors in ratio calculation that leads to a decision to delete is a mistake in the recording of information from the financial statement. As previously mentioned, financial statements in databases include a number of easily observable errors. It is likely that there are other errors of information recording that may have led to extreme outliers. With large samples these errors, or other true outliers that are not representative of the population, as a whole will distort ratio calculations and distribution. To eliminate these outliers a number of techniques may be used. For example, the data can be transformed using a Common (Natural) Logarithm transformation (Pinches, Mingo and Caruthers, 1973). Another approach is Trimming the Data which takes a percentage of observations off of each tail (Kennedy, Lakonishok and Shaw, 1992). A two-stage transformation can be used using the Gamma Probability Distribution approach to normalizing a sample (Frecka and Hopwood, 1983). To minimize the effect of outliers, the sample data can be placed into a Rank Order for analysis (Kennedy, Lakonishok and Shaw, 1992). More recently, it has been suggested that the Box-Cox Transformation be used to normalize data (Sudarsanam and Taffler, 1995). The data observations may also be Winsorized and this reduces the impact of outliers while retaining observations (Kennedy, Lakonishok and Shaw, 1992). Table 3 presents the strengths and weaknesses of each of the transformation techniques.

Table 3 About Here

Each of these techniques reduces the effect of outliers and brings the corrected data mean closer to the median with a reduction in the standard deviation. However, several of the techniques for transforming the data have a major drawback in that they are not defined for negative observations. This is a problem because some financial ratios have a significant portion of the observations below zero (for example, return ratios). Other techniques are not very effective at transforming the data so that normal distributions result. To be effective requires a shifting of the data points and this makes it possible to compare mean, median and standard deviation for data that has been Trimmed or Winsorized. Of these two techniques, Winsorized data at five percent of each tail brings the means and medians most closely together, reduces the standard deviation the most, and provides the good approximate value for all ratios. Table 4 provides the means, medians and standard deviations for the Winsorized sample data.

Table 4 About Here

Based on the comparison of each of the transformation techniques, it is recommended that the KCEL financial statement database be Winsorized using five percent of each tail. This process provides a good distribution with close means and medians and much lower standard deviations. While the effect of extreme outliers is reduced, the number of financial statements and data points is unchanged and provides a more accurate view of firm financial performance.

Table 5 shows the degree of the skewness of the distribution of the ratios along with the kurtosis (peaked ness) of the distribution for the data Winsorized at five percent of each tail. Using a common technique for calculating skewness[1], a symmetric distribution will have a value of zero, while a skew to the left will have a negative value (usually not more than three). If the negative or positive value is greater than three, the distribution is highly skewed and cannot approximate multivariate sample assumptions (it is desirable to be as close to zero as possible). For kurtosis, a value of 0 indicates a normal peak in the distribution (Mesokurtic), greater than 0 indicates a high peak (Leptokurtic) and a

[1] ***Formulas have not been included in this paper to save space. The calculations for each transformation technique, distribution and kurtosis are available by contacting the authors.***

value below 0 indicates a flat (Platykurtic) distribution (Foster, 1986; Glass and Stanley, 1970).

Table 5 About Here

Conclusion

When researchers use databases in general and specifically when they use the KCEL financial statement information, adjustments must be made if analysis is to be meaningful. As a first step, the sample selected must be carefully reviewed for obvious information entry errors. With the sample of wholesale firms described in this paper, 65 firm financial statements contained errors of such magnitude that they needed to be eliminated from study. This is a common problem with data and careful review is necessary to be certain that accurate observation measurements are being used in the analysis stage. If ratio variables are part of the analysis being undertaken, the remaining data must be reviewed to remove financial observations that lead to denominators being negative or approaching zero. Because ratio analysis assumes proportionality between denominators and numerators of observations being compared, this is an essential adjustment that must be made for results to have meaning. For the wholesale firm sample discussed in this paper, of the 185 firms reviewed for this adjustment, 19 firms were removed because their performance data resulted in ratio denominators that were negative or approaching zero. If these observations had been included in the data being analyzed, the results would have been distorted beyond meaning.

Even with the elimination of sample financial statements and adjustment for negative and approaching zero denominator removal, the sample may have extreme observations that distort mean and standard deviation calculations. As a result, the calculated means do not conform to needed assumptions that distributions approximate normal distribution for the use of multivariate analysis techniques. For this reason, the data must be transformed to achieve normality parameters if parametric statistics are to be used. Based on our findings, it is recommended that the corrected and adjusted data be Winsorized with five percent of each observation distribution tail being given the next extreme observation in the distribution.

If these steps are followed, data distributions meet sample assumptions with one major concern. That is, with the sample size moving from 250 firms to 166, concern is raised about how representative the sample is of the population. As an alternative to the removal of records containing obviously erroneous data points and if the representativeness and statistical power of the research data are of a serious concern, those particular data points can be smoothed using mean substitution. Regardless of which approach is taken, the fact that errors are removed and data transformed gives confidence that the results of analysis will be meaningful and those inferences may be drawn from the data.

In addition, if all researchers using large databases use the information adjustment techniques recommended in this paper, or use databases that have been previously adjusted, the result of analysis will be consistent and may be compared with the results obtained by different investigators. In any event, this paper presents recommended adjustments to be taken when working with financial statement data samples to help meet required characteristics for the use of valid multivariate statistical analysis.

Bibliography

Altman, I. E., R. G., Narayana, P., "Zeta Analysis: A New Model to identify Bankruptcy Risk of corporations," Journal of Banking and Finance, June 1977, 29-54.

Altman, I. E., Avery, R. B., Eisenbeis, R. A., and Sinkey, Applications of Classification Techniques in Business, Banking, and Finance, JAI Press, 1981.

Asch, L., "How the RMA/Fair Isaac Credit-Scoring Model was built," Journal of Commercial Lending, June 1995, 10-16.

Aziz, A., Emanuel, D. C., and Lawson, G. H., "Bankruptcy Prediction: An Investigation of Cash Flow Based Models," Journal of Management Studies, September 1988, 419-437.

Beaver, W. H., "Financial Ratios as Predictors of Failure," Journal of Accounting Research, supplement, 1966, 71-127.

Beedles, William L. and Simkowitz, Michael A., "A Note on Skewness and Data Errors," The Journal of Finance, Vol. XXXIII, No. 1, March 1978, 288-293.

Bennin, Robert, "Error Rates and COMPUSTAT: A Second Look," The Journal of Finance, Vol. XXXV, No. 5, December 1980, 1267-1271.

Blum, M., "Failing Company Discriminant Analysis," Journal of Accounting Research, Spring, 1974, 1-25.

Casey, C. J., & Bartczak, N. J., "Cash Flow: It's not the bottom line," Harvard Business Review, July/August 1984, 61-66.

Courtenay, S. M. and Keller, S. B., "Errors in Databases - An Examination of the CRISP Shares-Outstanding Data," Accounting Review, Vol. 69, Iss 1, 1994, 285-291.

Deakin, E., "A Discriminant Analysis of Predictors of Business Failure," Journal of Accounting Research, Spring, 1972, 167-179.

Dietrich, J. R., & Kaplan, R. S., "Empirical Analysis of the Commercial Loan Classification Decision," Accounting Review, January 1982, 18-38.

Duchessi, P., Shawky, H., Seigle, J., "Commercial Loan Analysis: An Application of Knowledge-Based Systems," Proceedings of the First Annual Conference on Expert Systems in Business, 1987, 89-98.

Duchessi, P., Shawky, H., Seigle, J., "A knowledge-engineered system for commercial loan decisions," Financial Management, Autumn, 1988, 57-65.

Fernandez-Castro, A., "Towards a General Non-parametric Model of Corporate Performance," Omega, International Journal of Management Science, Vol. 22, No. 3., 1994, 237-249.

Foster, George, Financial Statement Analysis, Prentice-Hall, Englewood Cliffs, 1986.

Frecka, T. J., and Hopwood, W. S., "The Effects of Outliers on the Cross-Sectional Distributional Properties of Financial Ratios." The Accounting Review, Vol. 58, 1983, 115-129

Gentry, J. A., Newbold, P., & Whitford, D. T., "Classifying Bankrupt Firms with Funds Flow Components," Journal of Accounting Research, Spring, 1985, 146-160.

Gentry, J. A., Newbold, P., & Whitford, D. T., "Funds Flow Components, Financial Ratios, and Bankruptcy," Journal of Business Finance and Accounting, Winter, 1987, 595-606.

Glass, Gene V. and Stanley, Julian C., Statistical Methods in Education and Psychology, Prentice-Hall, Englewood Cliffs, 1970.

Grice, John S. and Ingram, Robert W., "Tests of the generalizability of Altman's bankruptcy prediction model, Journal of Business Research, October 53-61, 2001

Guertin, Wilson H. and Bailey, John P. Jr., Introduction to Modern Factor Analysis, Edwards Brothers, Ann Arbor, 1966.

Hyndman, C., "Are bank regulators ready for credit scoring of commercial loans?" Commercial Lending Review, Winter, 1994, 92-95.

Johnson, Richard and Wichern, Dean W., Applied Multivariate Statistical Analysis, Prentice-Hall, Englewood Cliffs, 1982.

Kennedy, D., Lakonishok, J. and Shaw, W. H., "Accommodating Outliers and Nonlinearity in Decision Models; Professional Adaptation", Journal of Accounting, Auditing and Finance, Vol. 7, 1992, 161-194

Kim, Dongcheol, "A reexamination of firm size, book-to-market, and earnings price in the cross-section of expected stock returns," Journal of Financial and Quantitative Analysis, December 463-489, 1997.

Kinney, Michael R. and Swanson, Edward P., "The accuracy and adequacy of tax data in COMPUSTAT," The Journal of the American Taxation Association, Spring 121, 1993.

Klein, B. D., Goodhue, D. L. and Davis, G. B., "Can humans detect errors in data?" MIS Quarterly, Vol. 21, Iss 2, 1997, 169-194.

Largay, J. A., and Stickney, C. P., "Cash Flows, Ratio Analysis, and the W.T. Grant Company Bankruptcy," Financial Analysis Journal, July/August, 1980, 51-54.

Lau, A. H., "A Five State Financial Distress Prediction Model," Journal of Accounting Research, Spring, 1978, 127-138.

Malhorta, D., Malhorta, R., McLeoad, R. W., "Artificial Neural Systems in Commercial Lending", Bankers Magazine, Nov/Dec 1994, 40-44.

Marais, M., Patell, J., and Wolfson, M., "The Experimental Design of Classification Models: An Application of Recursive Partitioning and Bootstrapping to Commercial Loan Classification," Journal of Accounting Research, Supplement, 1984, 87-113.

Mensah, Y. M., "The Differentiated Bankruptcy Predictive Ability of Specific Price Level Adjustments: Some Empirical Evidence." The Accounting Review, 228-245.

Nenide, Boris, unpublished working paper, 2001.

Norton, C. L., & Smith, R. E., "A Comparison of General Price Level and Historical Cost Financial Statements in the Prediction of Bankruptcy," The Accounting Review, January, 1979, 72-87.

Ohlson, J. A., "Financial Ratios and the Probabilistic Prediction of Bankruptcy," Journal of Accounting Research, Spring, 1980, 109-131.

Pinches, G. E., Mingo, K. A. and Caruthers, K. J., "The Stability of Financial Patterns in Industrial Organizations," Journal of Finance, Vol. 28, June 1973, 369-396.

Pricer, R. & Johnson, A., "An Innovative Method for Accurately Predicting the Loan Payment Performance of Small Businesses," Proceedings: ICSB 41st World Conference, June 17-19, 1986, 359-387.

Rosenberg and Houglet, "Error Rates in CRISP and COMPUSTAT Data Bases and Their Implications," Journal of Finance, Vol. 29, September 1974.

Shah, Jasymeen R. and Murtaza, Mirza B., “A neural network based clustering for bankruptcy prediction,” American Business Review, June 80-86, 2000

Shaw, M. J., & Gentry, J. A., "Using an Expert System with Inductive Learning to evaluate Business Loans," Financial Management, Autumn, 1988, 45-56.

Stohr, E., "Decision Support and Knowledge-Based System," Journal of Management Information Systems, a special edition, December 1986.

Strischek, M., "RMA Lending Outlook for 1995," Business Credit, June 1995, 47-48.

Sudarsanam, P. S., and Taffler, R. J., “Financial Ratio Proportionality and Inter-Temporal Stability: An Empirical Analysis.” Journal of Banking and Finance, Vol. 19, 1995, 45-61

Zavgren, C. V., "Assessing the Vulnerability of American Industrial Firms: A Logistic Analysis," Journal of Business Finance and Accounting, Spring 1985, 19-45.

Table 1 ***Sample Mean, Median and Standard Deviation of Financial Ratios Before and After Data Error Removal***

Financial Ratio	Mean		Median		Standard Deviation	
	Before Error Removal (n = 250)	After Error Removal (n = 185)	Before Error Removal (n = 250)	After Error Removal (n = 185)	Before Error Removal (n = 250)	After Error Removal (n = 185)
Current Ratio	3.62	2.42	1.15	0.97	24.54	11.96
Quick Ratio	8.07	4.43	2.17	2.08	64.54	16.40
Current Liabilities to Net Worth	144.30	126.08	77.25	77.64	261.77	198.69
Total Liabilities to Inventory	172.23	149.54	98.30	101.74	216.73	179.50
Total Liabilities to Net Worth	177.95	166.91	105.45	105.45	415.54	214.33
Fixed Assets to Net Worth	49.14	44.49	19.30	20.74	112.49	111.08
Collection Period	42.20	41.42	36.90	37.80	47.33	41.33
Inventory Turnover	14.74	16.02	8.25	8.56	28.09	25.50
Assets to Net Sales	43.08	40.77	33.35	34.49	70.37	37.72
Net Sales to Net Working Capital	19.43	4.66	7.85	7.84	76.17	67.49
Accounts Payable to Net Sales	6.95	6.60	5.65	6.33	6.44	6.15
Return on Sales	2.62	1.38	1.20	1.28	11.83	9.49
Return on Assets	11.17	6.47	3.80	3.88	46.35	13.83

Return on Net Worth	22.84	14.71	8.75	8.99	93.87	27.49

Table 2 ***Sample Mean, Median and Standard Deviation of Financial Ratios Before and After Data Error Removal and Adjustment for Negative and Approaching Zero Denominators***

Financial Ratio	Mean		Median		Standard Deviation	
	Before Error Removal and Adjustment (n = 250)	After Error Removal and Adjustment (n = 166)	Before Error Removal and Adjustment (n = 250)	After Error Removal and Adjustment (n = 166)	Before Error Removal and Adjustment (n = 250)	After Error Removal and Adjustment (n = 166)
Current Ratio	3.62	2.58	1.15	1.00	24.54	10.52
Quick Ratio	8.07	4.31	2.17	2.01	64.54	15.15
Current Liabilities to Net Worth	144.30	113.79	77.25	68.35	261.77	138.94
Total Liabilities to Inventory	172.23	146.33	98.30	98.77	216.73	169.95
Total Liabilities to Net Worth	177.95	143.16	105.45	90.10	415.54	156.69
Fixed Assets to Net Worth	49.14	30.71	19.30	18.07	112.49	37.85
Collection Period	42.20	42.25	36.90	38.72	47.33	40.97
Inventory Turnover	14.74	14.87	8.25	8.28	28.09	25.32
Assets to Net Sales	43.08	40.19	33.35	34.84	70.37	28.43
Net Sales to Net Working Capital	19.43	8.79	7.85	7.34	76.17	15.54
Accounts Payable to Net Sales	6.95	6.17	5.65	6.35	6.44	6.66
Return on Sales	2.62	1.84	1.20	1.38	11.83	9.11
Return on Assets	11.17	6.53	3.80	3.91	46.35	12.20

Return on Net Worth	22.84	12.98	8.75	8.06	93.87	23.15

Table 3 ***Transformation Technique Strengths and Weaknesses (5 equals strong, 0 equals weak)***

	Provides Normal Distribution	Brings Mean Close to Median	Reduces Standard Deviation	Defined for Negative Observations	Able to Compare Mean, Median & Standard Deviation
Logarithm	3	2	4	0	0
Trimming	4	4	4	5	5
Gamma Probability	4	3	3	3	3
Rank Order	3	3	3	0	0
Box-Cox	3	3	3	0	0
Nenide Power	5	4	4	4	0
Square Root	3	3	3	0	0
Winsorize	5	5	5	5	5

Table 4 ***Sample Mean, Median and Standard Deviation of Financial Ratios Before and After 5% Winsorizing***

Financial Ratio	Mean		Median[2]		Standard Deviation	
	Before Winsorizing	After Winsorizing	Before Winsorizing	After Winsorizing	Before Winsorizing	After Winsorizing
Current Ratio	2.58	1.46	1.00	1.00	10.52	1.18
Quick Ratio	4.31	2.31	2.01	2.01	15.15	2.10
Current Liabilities to Net Worth	113.79	76.17	68.35	68.35	138.94	79.13
Total Liabilities to Inventory	146.33	103.03	98.77	98.77	169.95	75.78
Total Liabilities to Net Worth	143.16	93.66	90.10	90.10	156.69	99.65
Fixed Assets to Net Worth	30.71	26.90	18.07	18.07	37.85	26.80
Collection Period	42.25	38.01	38.72	38.72	40.97	17.39
Inventory Turnover	14.87	12.60	8.28	8.28	25.32	10.67
Assets to Net Sales	40.19	37.55	34.84	34.84	28.43	15.63
Net Sales to Net Working Capital	8.79	7.65	7.34	7.34	15.54	7.90
Accounts Payable to Net Sales	6.17	6.27	6.35	6.35	6.66	3.37
Return on Sales	1.84	1.39	1.38	1.38	9.11	3.17
Return on Assets	6.53	4.17	3.91	3.91	12.20	4.75

[2] While effective at bringing the sample mean closer to the median and thereby reducing the standard deviation, symmetrical (the same percentage at both tails) Winsorizing has no effect on the sample median.

Return on Net Worth	12.98	9.53	8.06	8.06	23.15	5.27

Table 5 ***Kurtosis (Peakedness) and Skewness of Distribution of Financial Ratio Distributions Before and After 5% Winsorizing***

Financial Ratio	Kurtosis		Skewness	
	Before Winsorizing	After Winsorizing	Before Winsorizing	After Winsorizing
Current Ratio	153.68	1.32	6.08	1.65
Quick Ratio	139.59	1.79	11.49	2.11
Current Liabilities to Net Worth	13.68	0.92	3.16	1.09
Total Liabilities to Inventory	76.01	0.96	3.96	1.17
Total Liabilities to Net Worth	8.48	0.25	2.50	0.97
Fixed Assets to Net Worth	123.89	0.43	10.47	1.02
Collection Period	85.74	-0.32	8.11	1.73
Inventory Turnover	57.06	1.89	6.81	1.96
Assets to Net Sales	88.19	0.65	8.28	0.70
Net Sales to Net Working Capital	44.41	1.28	9.94	1.31
Accounts Payable to Net Sales	55.06	-1.04	6.06	0.38
Return on Sales	94.54	-0.03	3.74	1.08

Return on Assets	35.67	-0.35	4.57	1.16
Return on Net Worth	13.35	1.10	3.34	1.37