An Analytic Hierarchy Process Approach to Assessing the Risk of Management Fraud

Ashutosh Deshmukh, (E-mail: avd1@psu.edu), Pennsylvania State University - Erie
Ido Millet, (E-mail: ixm7@psu.edu), Pennsylvania State University - Erie

Abstract

The purpose of this study is to demonstrate how the Analytical Hierarchy Process (AHP) can be employed to assess the risk of management fraud. The paper illustrates how the AHP can be used to measure red flags on a scale of varying intensities, to aggregate various red flags, and to derive a single measure of risk of management fraud. The significant red flags identified in past empirical research are used in this study for illustrative purposes and the AHP models are implemented in Expert Choice (1997) software. The AHP allows auditors to customize their approaches in assessing the risk of management fraud for each audit engagement. Auditors can use significant red flags identified in the earlier studies, and also add or delete certain red flags if the engagement warrants it. The AHP can be used as a viable auditing tool in situations where auditors perceive the need to exercise judgment and or to arrive at a group consensus.

Introduction

The responsibility of an independent auditor concerning management fraud remains a contentious issue. The recently issued Statement on Auditing Standard (SAS) No. 82 "Consideration of Fraud in a Financial statement Audit," requires the auditor to plan and perform the audit to provide a reasonable assurance that the financial statements are free of management fraud. The standard provides operational guidance in assessing the risk of material misstatements due to fraud and provides risk factors that the auditor should consider during the audit.

Readers with comments or questions are encouraged to contact the authors via e-mail.

Management fraud is associated with explosion of litigation against the auditor (Palmrose, 1991, Carcello and Palmrose, 1994). Costello (1991), after an excellent analysis of court cases involving management fraud, argues that neither generally accepted auditing standards (GAAS) nor SAS No. 53 constitute the controlling measures of an auditor's responsibility and advises that auditors should design their audits to detect all forms of fraud regardless of cause and regardless of whether suspicious circumstances are present. Auditors need to plan their audits after a careful assessment of the risk of management fraud.

Auditors traditionally have used personal, business, and economic risk factors (referred to as red flags) in assessing the risk of management fraud and planning the audit. Red flags are potential symptoms existing within a company's business environment that would indicate a higher risk of an intentional misstatement of financial statements (Price Waterhouse, 1985). SAS No. 82 cites several examples of red flags such as excessive interest by management to maintain or increase the entity's stock price or earnings trend through the use of unusually aggressive accounting practices, high turnover of management, counsel, or board members, and overly complex organizational structures without apparent business purpose.

The purpose of this paper is to demonstrate how the Analytical Hierarchy Process (AHP) can be employed to assess the risk of management fraud. The paper illustrates how the AHP can be used to measure red flags on a scale of varying intensities, to aggregate various red flags, and to derive a single measure of risk of management fraud. The earlier empirical research on red flags (Loebbecke, Eining and Willingham, 1989; Bell, Szykowny, and Willingham, 1993; Fanning, Cogger, and Srivastava, 1995; Hansen, McDonald, Messier, and Bell, 1996) has identified significant red flags that can be used on audit engagements. This has provided auditors with shorter lists of red flags; however, aggregation of red flags still needs to be done by cumbersome statistical or neural network methods, which may cause problems in deploying this methodology in the field. Also, each audit engagement is unique, and based on the industry, nature of business, or the experience or judgment of the auditor, different red flags may be deemed important. In such situations auditors are left with a checklist of red flags but no means of aggregating the red flags to derive a measure of risk of management fraud. The AHP approach can be used as a practical auditing tool in such situations to assess the risk of management fraud.

The rest of the paper proceeds as follows. First, we briefly review the empirical research of management fraud based on red flags and identify pertinent limitations of the research. Secondly, using the empirical results of the earlier research we introduce the AHP and after contrasting it with the popular scoring table technique, we develop an AHP model for the risk assessment of management fraud. Then we discuss the advantages and disadvantages of the AHP. Finally, we discuss the conclusions and extensions of this research.

Literature Review

This paper uses the model developed by Loebbecke and Willingham (1988) and Loebbecke, Eining, and Willingham (1989). Loebbecke and Willingham (1988) proposed a causal model to assess the likelihood of management fraud. The model was specified as follows:

$$P\,(MI) = f\,(C, M, A)$$

P (MI) is the probability of management fraud (or material irregularities), C is the degree to which conditions are such that a material management fraud can be committed, M is the degree to which person or persons in positions of authority and responsibility in the entity have a reason or motivation to commit management fraud, and A is the degree to which the person or persons in the positions of authority and responsibility in the entity have an attitude or set of ethical values such that they would allow themselves (or even seek) to commit management fraud, and where, if C, or M, or A = 0, then P(MI) = 0. Thus, this model posits that if one of the requirements C, M, or A is not present, then the likelihood of management fraud is zero.

Loebbecke and Willingham attempted to validate the model by categorizing red flags provided in SAS No. 53 according to the model's three components. Then the authors' conducted a frequency analysis of red flags in management fraud cases reported in Securities and Exchange Commission (SEC) Accounting and Auditing Releases (AAER). The authors refined the list of red flags, provided by SAS No. 53, by adding newly discovered red flags. Based on the frequency analysis, more frequent red flags were labeled as "primary indicators" and the other red flags were labeled as "secondary indicators." The modified model was reapplied to select AAER cases and the model was found to be a good fit.

Loebbecke, Eining, and Willingham (1989) extended the above research by conducting a survey of KPMG Peat Marwick partners who had encountered management fraud. The survey sought detailed information about management fraud incidents in terms of frequency of occurrence, nature, and impact on the client's financial statements. The survey resulted in an addition of new red flags, and deletion, reclassification, and recombination of some of the red flags in the model. The authors then presented a revised fraud assessment model that incorporated red flags, classified as "primary" and "secondary" indicators for the model's three components, as identified by the survey of the KPMG Peat Marwick partners and AAER cases. Bell, Szykowny, and Willingham (1993) enhanced this database by collecting data from engagements where no management fraud was discovered. The resultant red flags are presented below.

The red flags database collected by KPMG Peat Marwick was analyzed by various researchers. Bell, Szykowny, and Willingham (1993) tested this model using cascaded logit model estimation approach and other statistical techniques. Fanning, Cogger, and Srivastava (1995), Hansen, McDonald, Messier, and Bell (1996), Bell, Carcello, and Willingham (1996) analyzed the same data set using different methods. Fanning, Cogger, and Srivastava (1995) used artificial neural networks, Hansen, McDonald, Messier, and Bell (1996) used a generalized qualitative response model, and Bell, Carcello, and Willingham (1996) used a logistic regression

Table 1
Risk Assessment Model Applicable to Management Fraud

Adapted from Loebbecke, Eining, and Willingham (1989). P(MI) = f (C, M, A) where, **C** is the degree to which conditions are such that a material fraud can be committed, **M** is the degree to which person(s) in the positions of authority and responsibility in the entity have a reason or motivation to commit management fraud, and **A** is the degree to which person(s) in the positions of authority and responsibility in the entity have attitude or set of ethical values such that they would allow themselves (or even seek) to commit management fraud.

Primary indicators of C are:	Primary indicators of M are:	Primary indicators of A are:
- Dominated decisions - Major transactions - Related party - Weak internal control - Difficult-to-audit transactions	- Industry decline - Inadequate profits - Emphasis on earnings projections - Significant contractual commitments	- Dishonest management - Emphasis of earnings projections - Personality anomalies - Prior year irregularities - Lies or evasiveness - Aggressive attitude toward financial reporting
Secondary indicators of C are:	**Secondary indicators of M are:**	**Secondary indicators of A are:**
- Significant judgments - High management turnover - Decentralized organization - Assets subject to misappropriation - New client - Rapid growth - Inexperienced management - Conflict of interest	- Rapid growth - Rapid industry change - Sensitive operating results - Incentive compensation - Adverse legal circumstances - Significant portion of management's wealth - Management's job threatened	- Weak internal control - Conflict of interest - Poor reputation - Frequent disputes with auditor - Undue pressure on auditor - Disrespectful attitude

model. As shown in Table 2, the set of red flags identified as significant in each study is different; however, many significant red flags are common to all studies. The hit rates achieved by these studies ranged from 63% to 85% in fraud cases and 84% to 91% in no fraud cases. These results indicate that the statistical or neural network models can be used as viable auditing tools to assess the risk of management fraud. However, the statistical results of various studies indicate that the data set is complex and depending on the method of analysis, provides different red flags as significant.

Limitations of the research

The empirical research does a very good methodological job but the primary problems lie in the applicability of these approaches in the field. The statistical and neural network approaches require specialized expertise that may not be available in all auditing firms. The research also has not provided a standard set of red flags; depending on the method of analysis different red flags appear to be significant. Additionally, every auditing engagement is unique and the importance of red flags may vary according to the nature of each engagement.

The auditor is either forced to use a particular statistical method mechanically or evaluate different (and sometimes conflicting) sets of red flags. If the auditor decides that a particular red flag is important in a given situation but is not included in the empirical research discussed earlier then the auditor is left with a checklist of red flags with no means to aggregate those red flags. Thus, the earlier statistical approaches may create problems where the auditor's judgment differs from the significant red flags identified in the statistical research. The problems with using red flag checklists are already identified in Pincus (1989).

Another problem is with the measurement of red flags. The empirical research has measured red flags as present or absent (1 or 0). Consider a red flag "weak internal control." This red flag cannot be merely weak or not weak, there can be gradations such as weak, moderately weak, strong, or very strong. The earlier methodologies force auditors to use a binary approach to measure red flags where some information may be lost. The AHP approach allows measurement of red flags on an intensity scale, assignment of weights to red flags, and mathematical aggregation of the red flags to derive a single measure of risk.

Analytical Hierarchy Process

The Analytic Hierarchy Process, developed by Thomas L. Saaty (Saaty, 1977), addresses many of the problems mentioned above. It has been used extensively by government and business organizations. Expert Choice, Inc. (1997), one of the vendors of AHP software, which is used in this study, reports thousands of installations on PCs around the globe.

In selecting AHP to support actual fraud assessment we hoped to achieve several objectives. The methodology should be accurate and acceptable to those who use it. It should simplify the judgment process while being theoretically sound. It should also enhance our ability to examine and communicate our conclusions.

The following sections describe the limitations of the popular *scoring table* approach to multiple criteria decision making and demonstrate the advantages of the AHP through an illustrative fraud assessment model.

Limitations of Scoring Tables

The scoring table approach is a well-known method for combining rankings or scores on multiple criteria into a final tally. Each criterion is assigned a weight reflecting its relative importance, and each alternative is rated on a scale such as 1-to-5 on each criterion. Each alternative rating on each criterion is multiplied by the criterion weight and the sum of these products across all criteria provides the total score of each alternative.

Table 2
Comparison of Results - Significant Red Flags Found in Earlier Studies

Bell, Szykowny, and Willingham (1993)	Fanning, Cogger, and Srivastava (1995)	Hansen, McDonald, Messier, and Bell (1996)	Bell, Carcello, and Willingham (1996)
1. Dominated decisions 2. Weak Internal Control 3. Difficult-to-audit transactions 4. Contentious accounting issues 5. Decentralized organization 6. Rapid growth 7. Conflict of interest 8. Ownership (public vs. private) 9. Inadequate Profitability 10. Undue emphasis on earnings 11. Adverse legal circumstance 12. Solvency problems 13. Rapid growth 14. Collusion with outsiders 15. Need to cover up an illegal act 16. Management dishonesty 17. Undue emphasis on earnings 18. Management risk taking attitudes 19. Strong personality anomalies 20. Misstatements in prior periods 21. Management evasiveness 22. Aggressive attitudes 23. Disrespect for regulatory authorities	1. Does management place undue emphasis o meeting earnings projections or other quantitative targets? 2. Does management display significant disrespect for regulatory authorities? 3. Do client personnel display significant resentment of authority? 4. Do key managers exhibit strong personality anomalies? 5. Does the client have a weak internal control environment? 6. Are there frequent and significant difficult-to-audit transactions or balances? 7. Is the client confronted with adverse legal circumstances? 8. Is the direction of change in the client's industry declining with many business failures? 9. Does the substantial portion of management compensation depend on meeting quantified targets? 10. Is a significant portion of management's personal wealth in the form of holdings in the client entity? 11. Are key managers considered highly unreasonable?	1. Is there any need to cover up an illegal act? 2. Does your experience with management indicate a degree of dishonesty? 3. Is the client's organization decentralized without adequate monitoring? 4. Are there significant and difficult-to-audit transactions or balances? 5. Is a significant amount of judgment involved in determining the total of an account balance or class of transactions? 6. Does the client have solvency problems? 7. Does management display a propensity to take undue risks? 8. Is the client a public company? 9. Do client personnel display significant resentment of authority? 10. Does management display significant resentment of regulatory bodies? 11. Is this a new client? 12. Does a conflict of interest exist involving the client entity and/or its personnel? 13. Do key managers exhibit strong personality anomalies? 14. Have managers recently entered into collusion with outsiders? 15. Do accounting personnel exhibit inexperience or laxity in performing their duties? 16. Is management attitude toward financial reporting unduly aggressive? 17. Does management place undue emphasis o meeting earnings projections or other quantitative targets? 18. Is the client confronted with adverse legal circumstances?	1. Weak internal control environment 2. Rapid company growth 3. Inadequate or inconsistent relative profitability 4. Management places undue emphasis on meeting earning projections 5. Management lied to the auditors or was overly evasive 6. The ownership status (public vs. private) 7. Interaction term between a weak internal control environment and an aggressive attitude toward financial reporting

The scoring table provides two major advantages over past evaluation models. Since it is a simple and popular approach it faces less of a challenge on the user acceptance front. It also overcomes the all-or-nothing mode whereby red flags are assessed as either present or not present. Instead, it allows the auditor to provide an assessment of the degree to which the red flag is present. For example, if we use a scale of 1-to-5 then '1' can indicate that the red flag is not present at all and a '5' can indicate that the red flag is fully present.

While the scoring method addresses the need to impose structure and combine multiple criteria, it suffers from several limitations. On the technical side, it is simply wrong to treat evaluations on a Likert type scale such as 1-to-5 as real numbers that can be multiplied and summed. We know that '4' implies a higher presence of a red flag than a '2', but we don't know by how much.

Since the Likert type scale is not a ratio scale ('4' is not twice as much as '2') it cannot support operations such as multiplying by weights and summing. The process looks respectable only because we peg the gradations on the Likert scale to regular numbers. In actuality though, these numbers are nothing but labels.

This means that if two firms received scores of 400 and 450, we cannot say that the second firm is 12.5% more liable to experience accounting fraud. Even more distressing is the fact that we cannot even say that the firm with the higher score is more liable to experience accounting fraud. The final score is the result of a flawed numerical procedure that cannot be trusted.

On the modeling side, scoring tables provide little support to the process of organizing criteria and allocating weights to them. For example, if we limited ourselves to the 18 red flags with positive weights identified in the Hansen et al. Study (1996, pp. 1029) the scoring table approach would not allow us to simplify the picture by grouping similar red-flags into higher-level constructs. Since the human mind cannot cope with more than seven (plus or minus two) constructs at a time we should organize red flags in a hierarchy of criteria and sub-criteria.

Reflecting the need to organize complexity via a hierarchy, among other things, Loebbecke et al. (1989) has already grouped red flags into groups and sub groups. Such a structure can make it easier to discuss the model, adjust weights, and communicate the results to others. It can also reduce the risk of allocating too much weight to red flags because we fail to realize they overlap.

For example, the weight of the red flag "There is a need to cover up an illegal act" should probably not be evaluated in isolation from "Auditors experience with management indicates a degree of dishonesty" and "Officers of the company have entered into collusion with outsiders." Bell et al. (1993) have already grouped these red flags under the major criterion of "Attitude" and the sub-criterion of "Dishonest Management".

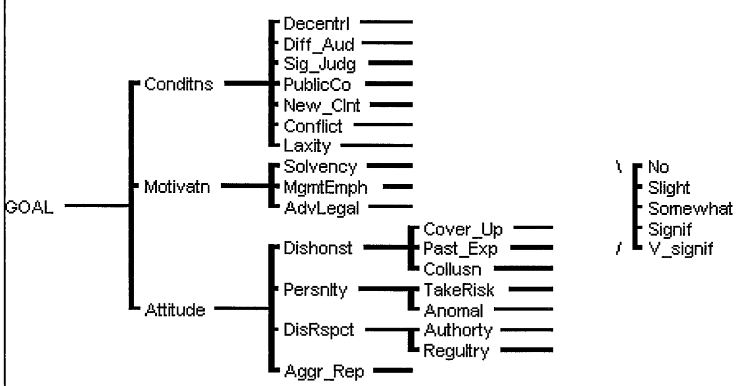
By structuring a hierarchical model of these red flags we can increase the manageability of the model and avoid inflating the weights of overlapping red flags. If we can also combine our evaluations using a ratio scale than we can overcome the major limitations of scoring tables. This is where the Analytic Hierarchy Process steps in.

A Hierarchy of Criteria

Several features of the AHP make it an extremely powerful method for handling complex multi-criteria decision-making problems. First, rather than treating criteria as a simple set, the AHP organizes them into a hierarchy. For example, Figure 1 shows how eighteen red flags can be organized into a three-level hierarchy. The hierarchical structure provides a natural mechanism for organizing and conquering complexity.

Figure 1: Construction of Hierarchy to Evaluate Risk of Management Fraud

Evaluate Risk of Management Fraud



Abbreviation	Definition
AdvLegal	Company is confronted with adverse legal circumstances
Aggr_Rep	Mgmt displays an aggressive attitude toward financial reporting
Anomal	Client personnel exhibit strong personality anomalies
Attitude	ATTITUDE (low ethics) that could allow or seek fraud
Authorty	Client personnel displays significant resentment of authority
Collusn	Company officers have entered into collusion with outsiders
Conditns	Degree to which CONDITIONS may allow material management fraud
Conflict	A conflict of interest within the company and/or its personnel
Cover_Up	There is a need to cover up an illegal act
Decentrl	Decentralized Organization
Diff_Aud	Difficult-to-Audit transactions
Dishonst	Dishonest management
DisRspct	Disrespecful Attitude
Dishonst	Dishonest management
Laxity	Accounting personnel exhibit inexperience or laxity
MgmtEmph	Management places undue emphasis on meeting earning projections
Motivatn	Degree of manager(s) MOTIVATION to commit fraud
New_Clnt	New client: no audit history, lack of info from previou auditor
No	This red flag is not present even to the slightest degree
Past_Exp	Auditor's experience with management indicates a degree of dishon

An Elicitation Process

The AHP also provides a powerful approach for evaluating the relative importance of criteria and the relative performance of alternatives under each final branch of the criteria hierarchy. The user can choose from a variety of graphical, numeric, or verbal elicitation methods for assigning relative weights to the elements under each node in the hierarchy. The responses of the user are converted using matrix algebra to relative weights for the compared elements. By breaking down the process to a series of simple pairwise comparisons and by extracting all the information in these pairwise comparisons, the cognitive strain is reduced, accuracy is increased, and inconsistencies can be detected and brought to the attention of the decision maker.

At the end of this elicitation process, we know the relative importance of each criterion and the relative performance of each alternative under each lowest-level sub-criterion. The weights assigned to each criterion node reflect its relative importance and the weights assigned to each alternative node reflects its relative performance under each lowest-level sub-criterion.

It is important to note that the AHP accommodates subjective as well as objective measurements. For example, the user may utilize a verbal or graphical elicitation process to express relative importance of the three top criteria in the hierarchy: Conditions, Motivation, and Attitude. Then, when asked to express the relative importance of the seven sub-criteria under the Conditions criterion, the user may enter the exact numerical coefficients derived by statistical analysis in past research. In both cases these measurements get converted into pure ratios within the AHP.

Local and Global Weights

Locally, at each node of the hierarchy, the relative weights of the sub-elements sum to one (1.0). This means that by multiplying the local weights found along a branch of the hierarchy, one can compute the global weight of a node.

In Figure 1, if the weight of the *Conditions* criterion is 22.8% and below it the *Decentralized* criterion captures a local weight of 20.1%, then the global weight of the *Decentralized* criterion will be 4.6% (22.8% * 20.1%).

Synthesizing Final Scores for the Alternatives

The alternatives capture a portion of the global weight of each lowest level node according to their relative performance on that sub-criterion. There are two synthesis methods within the AHP for doing this. The most common one is to let the best performing alternative under each sub-criterion capture the full global weight. The other alternatives then receive a fraction of that weight in proportion to their relative performance compared to the best alternative. The total score for each alternative is then computed as the sum of the weights it captured under all lowest-level sub-criteria.

Measurements on a Ratio Scale

The prominence of relative measurements and ratios within the AHP is another unique aspect of this methodology. Within the AHP all measurements are conducted on a ratio scale. Unlike *Utility Theory* where a final score of 40 utility points tells us nothing about the distance to a score of 20, in the AHP a score of 40% is precisely twice as much as a score of 20%. This makes the results of the process more useful and more amenable to further analysis.

Red Flag Analysis using the AHP

Given that Red Flag analysis of the potential for management fraud requires allocating weights to a significant number of criteria and combining subjective and objective observations, the AHP seems like a natural candidate. The following discussion illustrates how Red Flag analysis can be conducted using the AHP.

The model we shall develop here demonstrates how the AHP can accommodate situations where judgement about the performance of alternatives (or firms in our case) is conducted using an *absolute* rather than a *relative* measurement scheme. We shall not evaluate each firm by comparing its performance on each red flag to other firms. Instead we shall evaluate each firm under each red flag using a fixed intensity scale allowing the auditor to express the extent to which the red flag is present.

Constructing a Hierarchy of Criteria

Figure 1 depicts how the 18 red flags with positive weights identified in the Hansen et al. Study (1996, pp. 1029) can be structured into a hierarchy. Following the way Bell et al. (1993) have grouped red flags; *Conditions*, *Motivation*, and *Attitude* are the main criteria. The *Conditions* criterion serves to group 7 red flags while under the *Motivation* criterion there are only three significant red flags. The Attitude criterion leads to one direct red flag and three sub-criteria. The *Dishonesty* sub-criterion, for example, combines the three related red flags mentioned above.

Specifying Ratings for Lowest-Level Criteria

This particular type of an AHP model is known as a Rating model since for each lowest level criterion in the hierarchy we specify possible ratings. As shown in Figure 1, the rating intensities we elected to use for this model range from 'No', indicating this red flag is not present even to the slightest degree, to "V_Signif", indicating this red flag is present very significantly.

Figure 1 shows that we elected to use the same set of intensities for all lowest-level criteria. We could attach different rating scales to each sub criterion. For example, we could use a Yes or No ratings for clear cut red flags such as "Public Company" and behaviorally anchored ratings (BARS) for more subjective red flags such as "Client personnel exhibit strong personality anomalies." To simplify the discussion we elected to use a uniform rating scale for all red flags.

The AHP treats these rating intensities with great care. Instead of attaching arbitrary numeric labels, like scoring tables do, the AHP user goes through an elicitation process that establishes numerical values that reflect the subjective relative intensity attached to each rating. These intensities have a relative rather than absolute values. By convention, the AHP scales the results so that they sum to one.

As shown in Table 3, these relative numerical values can be strikingly different from those implied by the common 1-to-5 scoring table scale. For example, the intensity of the 'Significant' rating is 234% (0.241/0.103) more than the intensity of the 'Somewhat' rating. If we used a scoring table where Significant would map to '4' and 'Somewhat' would map to '3' then the ratio implied by the scoring table technique would have been 133% (4.0/3.0) which is quite different from 234%.

The methods used by the AHP to elicit and compute these numerical values are identical to those used for allocating weights to the criteria. They are described below.

Allocating Weights to Criteria and Ratings

The second step in applying the AHP involves allocating weights to each criterion and splitting that weight among the sub-criteria or Ratings below each criterion. A pairwise comparison process improves the accuracy of these weights since it allows auditors to focus on a series of relatively simple questions.

The accuracy of the process is bolstered

Table 3
Ratings Weights are Re-Scaled to Give Highest Rating a Weight of 1.0

Rating	Intensity	Re-scaled Weight
Very Significant	0.621	1.000
Significant	0.241	0.389 (0.241 / 0.621)
Somewhat	0.103	0.167 (0.103 / 0.621)
Slight	0.034	0.056 (0.034 / 0.621)
No	0.000	0.000 (0.000 / 0.621)

further by the fact that as the number of compared elements increases, the pairwise comparison process requires an increased level of redundancy. For example, evaluating the relative weights of the top three criteria in our model requires three pairwise comparisons, only one of which is redundant. By contrast, evaluating the relative weights of the seven red flags under the *Conditions* criterion requires 21 pairwise comparisons, 15 of which are redundant.

This redundancy also allows the methodology to raise an alarm when the user is inconsistent. For example, if we judged the *Conditions* and *Motivation* criteria to be equal in weight and the *Motivation* criterion to be twice as important as the *Attitude* criterion, then we should judge the *Conditions* criterion to be twice as important as *Attitude*. In short, transitivity should be observed.

Software implementations of the AHP, such as Expert Choice (1997), provide a variety of verbal, numeric, and graphical comparison methods via which the decision maker can express these pairwise comparisons. Below the surface, this input is converted into numerical ratios. Matrix algebra is then used to synthesize the relative weights of the compared elements.

For a faster though somewhat less accurate elicitation method (Millet, 1997), the decision-maker can express relative weights by directly adjusting the lengths of graphical bars. Another method of direct estimation is to provide actual numerical values in cases where weights or relative values are known. Figure 2 shows how the factor weights found in the Hansen et al. Study (1996, pp. 1029) were specified directly as the relative weights for the red flags under the *Conditions* criterion.

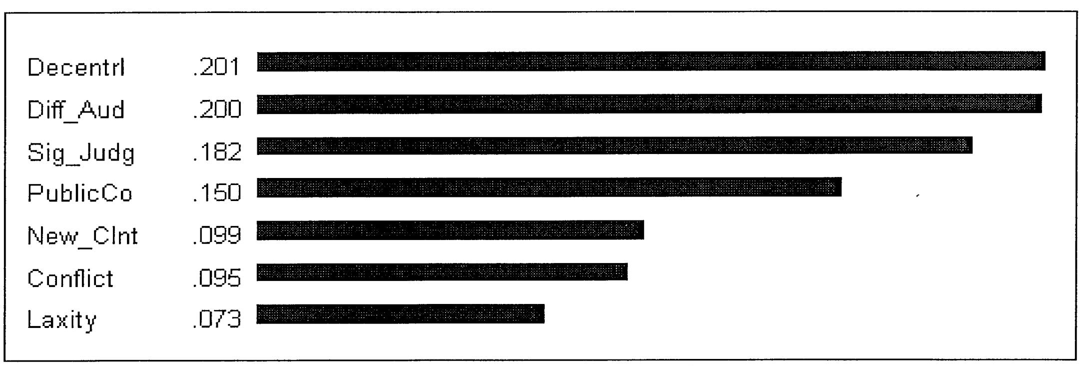
Figure 2: Specification of Weights for Red Flags

Evaluate Risk of Management Fraud

Node: 10000

Data with respect to: Conditns < GOAL

Decentrl	3.87
Diff_Aud	3.85
Sig_Judg	3.51
PublicCo	2.88
New_Clnt	1.9
Conflict	1.82
Laxity	1.41



Inconsistency Ratio =0.0

Splitting of Criteria Weights among Sub-criteria

The weight of each top criterion is split among its sub-criteria in proportion to their relative weights. This hierarchical allocation and splitting of weight addresses the problem of double counting. For example, as shown at the top of Figure 3, the weight of the *Attitude* criterion (71.8%) is split among four sub-criteria, one of which is *Dishonesty*. Similarly, the weight of *Dishonesty* criterion (59.6%) is split among three sub-criteria.

This hierarchical structure not only clarifies the structure of the problem but also provides a powerful method for adjusting weights. If we decide to shift weights among high-level criteria, the effect of these changes automatically cascades down the branches of the hierarchy. In other words, the user can reason and act at levels of abstraction that are higher and more intuitive than individual red flags.

Using the AHP Model

After the structuring and evaluation phases we are ready to use the AHP model for evaluating fraud risk. Figure 4 shows a portion of the evaluation spreadsheet generated by the AHP software. Each row in this spreadsheet contains the fraud potential evaluation for a single firm. The first two columns hold the name of the firm and its overall score. The rest of the columns hold the ratings the firm received on each red flag. Figure 4 shows only the first four red flags under the *Conditions* criterion. However, it contains columns for all 18 red flags in the lowest levels of the hierarchy tree (Figure 1).

In this spreadsheet the rating intensities are re-scaled so that the highest weighted rating

Figure 3: Splitting of Criteria Weights Among Sub-Criteria

Evaluate Risk of Management Fraud

Synthesis of Level 3 Nodes with respect to GOAL

Distributive Mode

LEVEL 1	LEVEL 2	LEVEL 3	LEVEL 4
Attitude=.718			
	Dishonst=.596		
		Cover_Up=.473	
		Past_Exp=.106	
		Collusn =.017	
	Persnlty=.055		
		TakeRisk=.035	
		Anomal =.020	
	DisRspct=.050		
		Authorty=.027	
		Regultry=.023	
	Aggr_Rep=.016		
Conditns=.228			
	Decentrl=.046		
	Diff_Aud=.046		
	Sig_Judg=.042		
	PublicCo=.034		
	New_Clnt=.022		
	Conflict=.022		
	Laxity =.017		
Motivatn=.054			
	Solvency=.038		
	MgmtEmph=.012		
	AdvLegal=.004		

(the "Very Significant" rating in our case) receives a weight of 1.00 and all other ratings receive a weight in proportion to that. Table 3 shows how this re-scaling is done.

The final score for each firm is calculated by multiplying the weight of each red flag (each column in the spreadsheet) by the weight of the rating it received and summing across all red flags. Since the weight of the highest rating ("Very Significant") is 1.0, and since the weights of all red flags sum to 1.0, a firm that scores "Very Significant" across all red flags would receive a maximum score of 1.000. As shown by Figure 4, rating four other firms across all red flags as "Significant", "Somewhat", "Slight", or "No" generates total scores of 0.389, 0.167, 0.056, and 0.000 respectively. These scores are simply a reflection of the weights of the re-scaled ratings as shown in Table 3 above.

Obviously, in most cases actual ratings are not going to be uniform. In such cases the weights of the red flags come into play. The last two rows in Figure 4 show total scores for two companies who both received 9 "Very Significant" evaluations and 9 "No" evaluations. However, since the first firm evaluated "Very Significant" on the nine lowest-weight flags its total score is only 0.136. In contrast, since the second firm evaluated "Very Significant" on the nine highest-weight flags its total score is a much higher 0.864.

Table 4 demonstrates how the total score of 0.026 was computed for the "Compute Example" row in the next to last row of the spreadsheet in Figure 4. To keep things simple, this row contains ratings only for three red flags. For a complete view of the computational procedure see Table 5, which shows how the total weight for "Sample Firm #9" was derived.

Advantages and Disadvantages of the AHP Approach

The AHP approach avoids the limitations of the scoring tables and is more flexible and user friendly when compared to statistically derived formulas that determine their weights from sample cases (Bell, Szykowny, and Willingham, 1993; Fanning, Cogger, and Srivastava, 1995; Hansen, McDonald, Messier, and Bell, 1996).

- User can adjust weights easily and intuitively either for individual red flags or for higher level criteria. This allows the AHP model to be adjusted to the specific realities

Figure 4: Evaluation Spreadsheet Generated by "Expert Choice"

	Alternatives	TOTAL	Conditns-Decentrl .0458	Diff_Aud .0455	Sig_Judg .0415	PublicCo .0341
1	All VERY SIGNIFICANT	1.000	V_signif	V_signif	V_signif	V_signif
2	ALL SIGNIFICANT	0.389	Signif	Signif	Signif	Signif
3	ALL "SOMEWHAT"	0.167	Somewhat	Somewhat	Somewhat	Somewhat
4	ALL "SLIGHT"	0.056	Slight	Slight	Slight	Slight
5	ALL "No"	0.000	No	No	No	No
6	"V_Signif on low weight"	0.136	No	No	No	No
7	"V_Signif on High weight"	0.864	V_signif	V_signif	V_signif	V_signif
8	Compute Example	0.026	Slight	Somewhat	Signif	
9	Sample Firm #9	0.142	Signif	V_signif	No	Slight
10						

Table 4: Computational Example of Total Score

Red Flag & Weight (see Figure 5)		Rating & Weight (see Table 3)		[Red Flag Weight] * [Rating Weight]
Decentrl	0.0458)	Slight	(0.056)	0.002
Diff_Aud	0.0455)	Somewhat	(0.167)	0.008
Sig_Judg	0.0415)	Signif	(0.389)	0.016
...15 other red flags not included				
Total Score:				0.026

of the engagement or to expert opinions within the auditing firm.

- The results can easily be communicated by simply looking at the model and tracing how total scores were derived.
- The software is easy to use.
- Ratio scale allows results to be compared meaningfully to various benchmarks such as total scores from past engagements. This can lead to the creation of standards such as "any firm with a total score above 0.4 should be examined at a high level of scrutiny."
- AHP is very tractable in modeling decision hierarchies. As the complexity and importance of decision increases deeper hierarchies can be used.
- AHP can accommodate group decisions. Auditors can collaborate in setting weights to red flags and the software supports the arranging of individual judgments into final weights in cases where consensus is not achieved by group members.

The AHP has been criticized for several reasons. There are theoretical and practical criticisms of the AHP, which have been briefly elaborated below.

- The original computational model of the AHP has been criticized for allowing the introduction of a new alternative, even an irrelevant one, to cause the rankings of the previous alternatives to be reversed (Belton and Gear, 1983). The emerging view is that the AHP should support two different computational modes: one which allows rank reversals and one which prevents it. Saaty (1994a, 1994b) directs us to select one of two computational modes: "distributive" or "ideal." The rating mode described in this paper prevents rank reversals.
- The subjective nature of the modeling process is a clear limitation. This means that the methodology cannot guarantee "correct" decisions or even agreement among multiple decision-makers. At best, it can only help our chances to make better decisions and to reach a consensus.
- Beyond the time and effort required to structure the AHP model, the number of pairwise comparisons increases rapidly as the number of nodes in hierarchy increases. Previous literature (Millet and Harker, 1990) has identified several ways for reducing the number of pairwise comparisons without significant reduction in accuracy. Additionally, the software implementation of the AHP offer alternative methods for eliciting judgments from decision-maker, which may reduce the time and effort. Still, the methodology requires more time and effort than less formal approaches.
- The AHP is only an aid, not a substitute, for analysis.

Conclusions

We demonstrate the use of AHP in assessing the risk of management fraud. AHP permits auditors to measure red flags on a scale of varying intensities and to aggregate red flags to derive a single measure of risk of management fraud. The significant red flags identified in past empirical research were used in this study for illustrative purposes, and the AHP models were

Table 5: A Computational Example using 18 Red Flags

The following table shows a detailed trace of how the total score is derived for Sample Firm #9 (shown as the 9th row in Figure 3).

Red Flags		Sample Firm #9		
		Rating		Score
Name	Weight	Name	Weight	[Red Flag Weight] * [Rating Weight]
Cover_Up	0.473	No	0.000	0.473*0.000= 0.000
Past_Exp	0.106	Slight	0.056	0.106*0.056= 0.006
Collusn	0.017	Somewhat	0.167	0.017*0.167= 0.003
TakeRisk	0.035	Signif	0.389	0.035*0.389= 0.014
Anomal	0.020	V_Signif	1.000	0.020*1.000= 0.020
Authorty	0.023	No	0.000	0.027*0.000= 0.000
Regultry	0.023	Slight	0.056	0.023*0.056= 0.001
Aggr_Rep	0.016	Somewhat	0.167	0.016*0.167= 0.003
Decentrl	0.046	Signif	0.389	0.046*0.389= 0.018
Diff_Aud	0.046	V_Signif	1.000	0.046*1.000= 0.046
Sig_Judg	0.042	No	0.000	0.042*0.000= 0.000
PublicCo	0.034	Slight	0.056	0.034*0.056= 0.002
New_Clnt	0.022	Somewhat	0.167	0.022*0.167= 0.004
Conflict	0.022	Signif	0.389	0.022*0.389= 0.009
Laxity	0.017	V_Signif	1.000	0.017*1.000= 0.017
Solvency	0.038	No	0.000	0.038*0.000= 0.000
MgmtEmph	0.012	Slight	0.056	0.012*0.056= 0.001
AdvLegal	0.004	Somewhat	0.167	0.004*0.167= 0.001
Total:				0.142

implemented in Expert Choice (1997) software. This approach allows auditors to customize their models for assessing the risk of management fraud for each audit engagement. Auditors can use significant red flags identified in the earlier studies, and also add or delete certain red flags if the engagement warrants it. The AHP can be used as a viable auditing tool in situations where auditors perceive the need to exercise judgment and or incorporate group input into a decision. We also discussed the advantages and disadvantages of the AHP.

Suggestions for Future Research

The use of AHP for assessing the risk of management fraud opens the door to several research questions. One line of empirical research could investigate if different auditors use different model structures or if within the same model structures different auditors assign different weights to different red flags. To the extent possible, we should use the results of this type of research in order to suggest template models for typical engagement situations.

Another research question is to what extent different auditors are consistent in rating the same firm using the intensity scales designed for each red flag. If, for example, one auditor rates a company as "Very Significant" on the Laxity red flag and another auditor rates the same company as "Very Slight" on that same red flag then our ability to deliver a reliable evaluation must be brought to question. On the positive side, such inconsistencies can be used internally by the auditing firm as a tool for training and for quality control.

Since the AHP models can be structured at various levels of detail, we should investigate the extent to which the additional effort required to evaluate more complex models is justified. For example, the illustrative model in Figure 2 breaks the personality criterion into two sub-criteria: Risk Taking and Personality Anomalies. If evaluating the Personality criterion directly generates very similar results to those achieved through the detailed evaluation of its sub-criteria then we may recommend simplifying the model.

The AHP approach to risk assessment can be compared to the statistical and neural net methods discussed in past literature. By applying these methods to the same cases, such comparisons can shed light not only on the relative ease of use and user acceptance of these methods but also on their relative accuracy.

An important research question that is not addressed by this paper is the mapping of final risk scores to appropriate action by the auditors. Obviously, as the level of risk score goes up, so should the level of scrutiny. The correct method for designing such audit policies should probably take into account the accumulation of experience within the auditing firm across all prior engagements and in particular those with the same audited client. 🕮

References

1. Auditing Standards Board, "Consideration of Fraud in a Financial Statement Audit," Statement of Auditing Standards, No. 82, New York, AICPA, Auditing Standards Board, 1997.
2. Auditing Standards Board, "The Auditor's Responsibility to Detect and Report Errors and Irregularities," Statement of Auditing Standards, No. 53, New York: AICPA, Auditing Standards Board, 1989.
3. Bell, T., J. Carcello, and J. Willingham, "Developing a decision aid for assessing the likelihood of fraudulent financial reporting," Working paper, 1996.
4. Bell, T., S. Szykowny, and J. Willingham, "Development of a Decision Aid for Assessing the Likelihood of Fraudulent Financial Reporting," Working paper, KPMG Peat Marwick, 1993.
5. Belton, V. and T. Gear, "On a Shortcoming of Saaty's Method of Analytic Hierarchies," *Omega*, December, pp. 228-230, 1983.
6. Carcello, J. and Z. Palmrose, "Auditor litigation and modified reporting on bankrupt clients," *Journal of Accounting Research*, 32, Supplement, pp. 1-30, 1994.
7. Costello, J., "The Auditor's Responsibilities for Fraud Detection and Disclosure: Do the Auditing Standards Provide Safe Harbor?" *Maine Law Review*, Vol. 43, pp. 265-305, 1991.
8. Expert Choice, Version 9, Expert Choice, Inc., Pittsburgh, PA, 1997. (*www. expertchoice.com*)
9. Fanning, K., K. Cogger, and R. Srivastava, "Detection of Management Fraud: A Neural Network Approach," *Intelligent Systems in Accounting, Finance, and Management*, Vol. 4, pp. 113-26, 1995.
10. Hansen, J., J. McDonald, W. Messier, and T. Bell, "A Generalized Qualitative-Response Model and the Analysis of Management Fraud," *Management Science*, Vol. 42, No. 7, pp. 1022-32, 1996.
11. Loebbecke J. and J. Willingham, "*Review of SEC Accounting and Auditing Enforcement Releases*," Working paper, KPMG Peat Marwick, 1988.
12. Loebbecke J., M. Eining, and J. Willingham, "Auditor's Experience with Material Irregularities: Frequency, Nature, and Detectability," *Auditing: A Journal of Practice & Theory*, Vol. 9, pp. 1-28, 1989.
13. Millet I. and P. Harker, "Globally Effective Questioning in the Analytic Hierarchy

Process," *European Journal of Operations Research*, Vol. 48, September, pp. 88-98, 1990.

14. Millet I., "The Effectiveness of Alternative Preference Elicitation Methods in the AHP," *Journal of Multi-Criteria Decision Analysis*, Vol. 6, No. 1, January, pp. 41-51, 1997.
15. Palmrose Z., "Trial of Legal Disputes Involving Independent Auditors: Some Empirical Evidence," *Journal of Accounting Research*, Vol. 29, pp. 149-185, 1991.
16. Pincus K. V., "The Efficacy of a Red Flags Questionnaire for Assessing the Possibility of Fraud," *Accounting, Organizations and Society*, Vol. 14, pp. 153-63, 1989.
17. Price Waterhouse, Challenges and Opportunity for the Accounting Profession: Strengthening the Public's Confidence, Price Waterhouse, 1985.
18. Saaty L. T, "A Scaling Method for Priorities in Hierarchical Structures," *Journal of Mathematical Psychology*, Vol. 15, pp. 59-62, 1977.
19. Saaty L. T., Fundamentals of Decision Making and Priority Theory with the Analytic Hierarchy Process, RWS Publications, Pittsburgh, PA, 1994a.
20. Saaty L. T., "How to Make a Decision: The Analytic Hierarchy Process," *Interfaces*, Vol. 24, pp. 19-43, 1994b.