UNIVERSITY OF SOUTHAMPTON

The Auditor's application of Analytical Procedures, the extent of their use and the effectiveness of such Procedures

by

Robin Napier Swan

Thesis submitted for the degree of Doctor of Philosophy

Faculty of Social Sciences

Department of Management

This thesis was submitted for examination in September 2001

This thesis was submitted for examination in September 2001.

This copy of the thesis contains the amendments suggested by the examiners.

This thesis is the result of work done while in registered postgraduate candidature.

	Correction Sheet		
	Errata	**Page Number**	

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SOCIAL SCIENCES

DEPARTMENT OF MANAGEMENT

Doctor of Philosophy

THE AUDITOR'S APPLICATION OF ANALYTICAL PROCEDURES: THE EXTENT OF THEIR USE AND THE EFFECTIVENESS OF SUCH PROCEDURES

by Robin Napier Swan

The audit profession is under great pressure to provide a service at the lowest price consistent with their professional standards. Burton and Fairford (1982) observed: Pressures on prices should lead to greater audit efficiency as firms strive to reduce costs to remain competitive. Thus audit efficiency will likely occur primarily at the level of hours committed to an engagement. These are hours that can be best replaced by the increased use of analytical techniques (now known as Analytical Procedures) and modern computer technology. This prediction has been proved correct. Analytical Procedures (APs) are now widely used; for example, the time devoted to them is claimed to be at least 33% of an audit engagement.

Auditors regard the procedures as very powerful and, since 1982, have consistently claimed that over 40% of all errors are initially identified by the application of an AP. Despite this, doubts have been raised about the effectiveness of APs. For example: our findings appear to support Blocher and Willingham and others that analytical review may be good at spotting the presence of errors, but does not reliably indicate the absence of errors (Loebbecke and Steinbart (1987)).

This thesis examines the environment in which APs are used. It considers surveys about the application of APs, including the extent of use of such procedures and explores whether APs are reliable, effective and efficient. An AP is defined as effective if both false-positive and false-negative indications of material error have low probabilities.

It is shown that APs are an important audit tool at all stages of the audit process, including as substantive procedures. APs are used extensively and simple techniques are the most common. Auditors perceive that APs are effective. New experiments show that certain APs have high frequencies of false signals. Our investigations imply that APs may be low cost but, in certain circumstances, they are ineffective, rather than cost effective.

List of Contents

List of Tables and Figures

Dedication

To my Mother, who always wanted the best for her boys.

Acknowledgements

I would like to acknowledge the support given by the National Audit Office while I was in their employ. I am particularly grateful to Sir John Bourn for his suggestions and support.

I would like to thank my supervisors Professor Michael Bourn, Dr Cecilio Mar-Molinero and particularly Professor Fred Smith. All contributed support, wisdom and encouragement.

Thanks also to many friends who would not let me give up. Particularly to Louise Swan, Gill Boulter and Tony Webster from Wokingham and to John Haworth, John McEwen, Steve Ardron and Martin Sinclair of the National Audit Office.

Abbreviations used

[The page number is where it first appears]

Abbreviation	Title or Name in full	Page No
AAA	American Accounting Association	14
AICPA	American Institute of Certified Public Accountants	25
APB	Auditing Practices Board	9
APC	Auditing Practices Committee	222
APs	Analytical Procedures	4
AR	Audit Risk	20
Ar	Analytical Review Risk	24
ARIMA	Autoregressive integrated moving-average model	165
ARM	Audit Risk Model	20
BRAM	Business Risk Audit Methodology	2
CR	Control Risk	20
DR	Detection Risk	21
DTI	Department of Trade and Industry	9
DVLA	Driver and Vehicle Licensing Agency	52
E	The error level seeded in the experiments	197
EDA	Exploratory Data Analysis	173
ICAEW	Institute of Charter Accountants of England and Wales	6
IDA	Initial Data Analysis	47
IR	Inherent Risk	20
JMU	Joint Monitoring Unit	82
KPMG	KPMG Peat Marwick	2
LR	Land Registry	190
NAO	National Audit Office	5
PGO	Paymaster Generals Office	190
PWC	PricewaterhouseCoopers	2
p_E	The frequency of observing $\lvert Y_T - Y_B \rvert \leq C$	200
p_O	The frequency of observing $\lvert \hat{Y}_T - Y_B \rvert \leq C$	200
SAS	Statements of Auditing Standards	5

SR	Sampling Risk	24
STAR	Statistical Techniques for Analytical Review	79
X11	Model for seasonal adjustment of times series	165
Y_B	The actual book value	8
Y_T	The true book value	59
$\hat{Y}_T$	Estimate of true value based on audit data	8

Chapter 1: Are Analytical Procedures a problem for the auditor?

1.1 Introduction

Auditing depends on trust, and as one audit partner has stated: '*If auditors lose that trust they have lost everything*[1]'. Recent events have focussed the public spotlight on the audit profession in a way that has called that trust into question. For example there have been a series of serious errors of judgement by auditors in their work with major companies such as Enron, WorldCom, Sunbeam, Cendant Corporation etc. These have led to accounting problems and even company collapse. The summer of 2002 saw a climax to this phenomenon with the disintegration of one major audit firm: Arthur Anderson.

As well as this issue of a lack of trust there are different perceptions of what auditing provides. In his article on this subject, Humphrey (1997) identified a role-perception gap where there are differing views of what is expected from an audit between the auditor and a variety of the users of the audited financial statements. Those users can include shareholders, Parliament and the general public. This is sometimes known as the expectation-gap. Humphrey (1997) provided a definition of the expectation-gap, it was: *'a representation of the feeling that auditors are performing in a manner at variance with the beliefs and the desires of those for whose benefit the audit is being carried out'*. In the light of recent developments, such as those referred to in the first paragraph, some might say the expectation-gap has widened or, at least, become more apparent.

Financial Audit is not a static process; rather it is continually evolving. The reasons are many. They include attempting to meet changing requirements such as the expectation-gap, or an attempt to orientate the audit to the needs of the management of the audited organisation. Another reason may be to provide a particular audit firm with an audit methodology that they perceive is to their advantage in the marketplace.

The emphasis of the most recent developments in audit methodology is concerned with achieving a better understanding of the audited organisation's business and its related risks. This is accomplished by a structured approach to risk assessment with an

[1] From – 'Auditors can restore trust with professional credentials': an article by a partner from the audit firm BDO Stoy Hayward and printed in the Times: Business Section page 27 (1 August 2002).

evaluation of those risks, plus acquiring an understanding of how the management of the organisation addresses those risks through their own management strategies and systems of internal control, together with a subsequent evaluation of the residual risks to the auditor. This approach is known as the Business Risk Audit Model (BRAM). It is variously known as Business Measurement Process (KPMG), Business Audit (Arthur Anderson), Audit Innovation (Ernst and Young), PricewaterhouseCoopers Audit Approach (PWC) and Audit21 (National Audit Office). Other audit firms have their own brand names. At least one of these approaches was in place in 1997.

A feature of all of these approaches is a reduced emphasis on substantive tests and, in particular, the time devoted to such tests. Eilifsen, Knechel and Wallage (2001) while identifying '*that the new methods are designed to improve the uses of high-level competitive and strategic information during the audit*' pointed out that '*to the extent that traditional substantive testing is decreased or eliminated, an auditor runs the risk of overlooking errors that might have been detected under those approaches*'. They also counselled some caution about the new approach: '*There is some understanding of the limitations of traditional methods due to previous research; a similar level of experience and insight must be attained before we can similarly assess the new audit methods*'. In the light of the reductions in the use of substantive tests, or in the extent of such tests, and the unproven nature of the methodology of BRAM, it is important that the procedures used as substantive tests are effective and do not provide significant levels of false signals to the auditor.

On the 16 April 2002, the Times newspaper ran a feature that highlighted a number of outstanding court cases against the large UK audit firms – the big Five as they were then known. For example, it was reported that PricewaterhouseCoopers (PWC) '*Faces a series of legal actions in respect of its auditing of Gazprom*'. According to the Times newspaper, Gazprom is the world's largest gas company. Similar examples were provided for KPMG, Deloitte and Touche, Ernst and Young and Arthur Anderson. These examples may, or may not, be the result of the revised methodologies. But a shortcoming, or failure, of any widely used audit procedure (test) could have additional adverse repercussions for the audit profession. This is particularly so now that there is a wider, more sceptical, audience for the work of the audit profession.

1.2 Outcomes of an audit test

The outcomes from the application of a substantive audit test can be divided into four categories according to (i) whether significant error is or is not present, and (ii) whether the audit test signals that significant error is, or is not, present. These four categories are shown in Figure 1.1 and lead to three different types of signal. For this thesis the signals are called correct, false-negative or false-positive. Such false signals are equivalent to type I and type II errors that are associated with statistical tests.

For the auditor, the consequences of a false-negative or a false-positive signal are very different. As a result of a test signalling significant error, the auditor should carry out further work to obtain evidence that either confirms the result indicated by the test, or establishes that there is no significant error. If it turns out that there was no error in the financial statements then the consequence is an economic one for the auditor. They have carried out additional, probably expensive, audit work because of a false-negative signal.

		True Position	
		There is no significant error	**There is significant error**
Audit Result	**The audit test signals that there is no significant error**	Correct signal	False-positive signal
	The audit test signals that there is significant error	False-negative signal	Correct signal

Figure 1.1: Possible outcomes from the application of a test. [Source: This thesis].

A false-positive signal could have much more serious consequences: it may cause, or be a factor in causing, the auditor to wrongly accept a set of financial statements. While a particular test is only one of many signals received by the auditor, it may be a key signal, and so any failure might have consequences for the performance of the audit. For example a substantive test might be such a key test, performed as a result of a number of judgments by the auditor. Such judgements may include an assessment of the risk to the business of its systems failing, other matters such as the business environment, and the consequent residual risks to the auditor. The substantive test may be a key source of evidence with the role of confirming or challenging the prior assumptions and judgements.

A false-positive signal may also have consequences for those who rely on the auditor's opinion. For example the users of the financial statements may suffer financial loss as a result of taking action based on the auditor's incorrect opinion. For the auditor their professional integrity is at risk, and they may also incur financial loss in the form of the payment of damages to those relying on his/her attestation of the financial statements. Examples of such losses include a settlement by Coopers and Lybrand of $95m in an out-of-court settlement over its work as auditor to MiniScribe, a bankrupt hard disk manufacturer. The Accountancy magazine for December 1992 reported that this settlement came months after the firm paid between $40m and $50m in a separate settlement over the affair. More recently the High Court in the UK has stated: '*that an auditor who, because of negligence, fails to spot that a company is insolvent, with the result that the company, with the auditor's approval, pays a dividend when it should not have done so, could be liable to the company for the amount paid out*' (Sasea Finance Ltd [in liquidation] v KPMG [a firm], The Times, 25 August 1998). These cases, and those described earlier, illustrate that the consequences of an incorrect opinion can be serious for an auditor: both to their reputation and financially.

1.3 Analytical Procedures

The subject of this thesis is Analytical Procedures (APs). These are audit procedures that are about the analysis of relationships. To carry out this analysis auditors collect, analyse and interpret data. The Glossary of Terms in the Auditing Standards define Analytical Procedures '*as the analysis of relationships:*

(a) Between items of financial data, or between items of financial and non-financial data, deriving from the same period; or

(b) Between comparable financial information deriving from different periods, to identify consistencies and predicted patterns or significant fluctuations and unexpected relationships, and the results of investigations thereof'.

This definition is entirely congruent with the definition of statistics given by Chatfield (1995): '*Statistics is concerned with collecting, analysing and interpreting data in the best possible way, where the meaning 'best' depends on the particular circumstances*'. The concepts and structures of a statistical approach are appropriate for APs. This does not necessarily, nor is it likely to, involve the application of sophisticated statistical

procedures such as Box-Jenkins time-series models. But what the two approaches should have in common is a consistent, structured strategy to the analysis. For a statistician, that approach is to formulate the problem, collect appropriate data, assess its structure and quality, carry out an initial examination of the data, select an appropriate analysis, interpret and communicate the result. Such a strategy would be consistent with the audit process, but is more explicit than is set out in the Auditing Standards, otherwise known as the 'Statements of Auditing Standards' (SASs). For example the Auditing Standards require the auditor to consider things such as data quality. In the SAS on APs, the auditor is required to: *'enquire of management as to the availability and reliability of information needed to apply analytical procedures and the results of any such procedures performed by the entity' (SAS 410).* The part underlined is, providing the auditor uses a critical analysis of the responses, about assessing the quality of the data. All the relevant factors that a statistician would expect are in paragraphs 12 to 16 of SAS 410, but nowhere does the SAS provide, or invite the consideration of, a structure to ensure that all the factors are explicitly considered. The components are there, but the SASs appear to be drafted for the auditor to interpret. This might be said to be a characteristic of all the SASs.

1.4 Why I did this research

Until I retired in October 2000, I was Director of Research and Statistics at the National Audit Office (NAO). Through my interaction with the audit firms, I saw an increasing application of Analytical Procedures (APs) by those firms, both as a planning tool and a substantive tool. I wanted to find out more: to say something about APs in their substantive role and whether those APs are effective. For this thesis, an effective audit procedure is taken to be one that does not provide high frequencies of false signals.

The question of the level of false signals is important since the audit approach relies heavily on judgements about the management of the organization being audited and the systems of controls that are in place. One question resulting from this is: when carrying out a substantive test, do auditors collect sufficient evidence to contradict, when that is appropriate, their judgements on the systems and the management? In particular if auditors use APs: are those procedures capable of doing that? So a further question is: what are the levels of false signals given by APs, particularly false-positive signals?

1.5 Research Methodology

The concern about the potential for false signals is only of interest if APs are an important source of audit evidence. This will be established in the thesis by two approaches: a consideration of the perceptions of auditors about the use of APs and a series of experiments to investigate the effectiveness of APs.

First, the thesis reviews a number of surveys which have considered the use of APs and which have been published in the literature. Then the thesis reports a survey based on a series of interviews with the technical partners in the largest sixteen audit firms in the UK. The interviews were structured conversations with the aim of gathering information to illustrate the extent of the use of APs, how auditors perceive the effectiveness of such procedures, their role in the audit process, and what quality checks the auditors require of the data, or of the structure of the APs etc.

It is established that APs are an important audit tool, both in terms of the extent of their use and their perceived effectiveness. This is a finding that is reinforced by a series of seminars run by the Audit Faculty of the Institute of Chartered Accountants of England and Wales (ICAEW) in 1998. They were called 'Tomorrow's Audit Today' and the theme was that the audit should be economic and add value for the client. A key theme of the seminars, that was re-enforced in the official session notes, was the *'elimination of transaction testing and their replacement by Analytical Procedures'*. Although this was not a formal statement of policy by the ICAEW, it was clearly an active encouragement from a leading professional body to increase the use of APs.

In the light of the importance of APs in the overall audit process, it is appropriate to establish whether APs are as effective as they are perceived to be by auditors. This thesis takes the definition of effective, "*producing the intended result*[2]", to mean that the AP will correctly signal the presence or absence of significant error. So the thesis will address the problem of the potential levels of false signals. This is first considered through a review of the literature that reported the results of a number of different experiments into the use of APs and the levels of type I and type II errors that were observed (false-negative and false-positive signals as defined in the thesis). Then the levels of false signals are explored through a series of experiments that use three

[2] Concise Oxford Dictionary: Tenth Edition

distinct sets of data as a basis for computer simulations. Two sets are based on pay data from different government agencies. These data were chosen because they are well controlled: they are of high quality and without large fluctuations. In a sense they are the ideal data for a procedure that explores relationships, such as APs. For example, if the AP produced high levels of false signals with such data, then any data set that is from a less well-controlled area, and/or data that is subject to more fluctuations, is likely to produce even higher levels of false signals.

The third set of data was created to be similar to that used by Knechel (1988a and 1988b). Knechel (and others) reported that APs were effective; their experiments used the reduction of other substantive tests as the criteria for effectiveness. One purpose of the new experiments was to investigate whether, if the criterion of test reduction had not been used, there would have been a different conclusion. The new experiments therefore measured directly the levels of false signals, using data and APs similar to Knechel's.

The thesis concludes by considering the implications of its findings, with particular reference to the use of the Business Risk Audit Model approach to audit.

1.6 The contribution to the knowledge of Analytical Procedures

This thesis contributes to the fund of knowledge about the application of APs: it introduces some new concepts and information about the use of APs and their effectiveness.

As part of the research, the idea was developed of false-negative and false-positive signals for APs. These were discussed on page 3 of this Chapter. The idea of false-negative and false-positive signals derives from the social sciences, but has not been used in the audit arena. In part, the use of this terminology was to overcome the misconceptions of some people that false signals (type I and type II errors) are only associated with statistical procedures. The concept of false-negative and false-positive signals was well received by a group of Directors from the National Audit Office at a presentation given to them in August 2000 as part of the development of ideas for this thesis. The concept allows the auditor to focus on the potential for an AP to give false signals, without a concern that such values can only be formally evaluated for statistical

procedures. It also makes clear that all estimates, or estimation processes, can potentially involve false signals.

A new survey, which is reported in Chapter 6, extends the scope of a number of earlier surveys (summarised in Chapter 5) by additionally investigating aspects of quality. These include data quality, data checking, the documentation of the AP process, measuring the closeness of the predicted value to the value being audited, and exploring whether and how the auditors validate the differences between those values. Issues about the extent to which APs replace other substantive tests are also explored, as are the auditors concerns about the application of some APs

Some earlier experiments, which are reported in Chapter 7, used the reduction of other substantive tests as the criterion for the effectiveness of APs. It is argued that the rationale they used is flawed. New computer simulations were carried out; using similar data and APs, to demonstrate that if the potential for false signals is measured directly, then the earlier experiments did not demonstrate the effectiveness of the APs. Then, using real data as a basis for the simulations, these findings are then extended into other types of AP to show the levels of false-negative and false-positive signals.

Finally, in order to provide some focus for some of the deficiencies in APs, a structure for the AP process is proposed. It involves disaggregating the AP process into three rules that enables the auditor to focus on the discrete stages of an AP, however simple or complicated it is.

The initial stage is to develop an expectation-value for the figure in the financial statements. This will be called the 'investigation-rule'. It seeks to incorporate causal links between different sets of data and utilises the auditor's understanding of the plausible relationships when using procedures such as ratio, trend or modelling analysis. It may be expressed as a sentence or as a formula.

The next stage is concerned with assessing the closeness of the expectation-value ($\hat{Y}_T$) to the book value (Y_B) that is being audited. The rule for assessing this closeness will be called an outcome-rule. The criterion that is associated with this rule, to help in the judgment of whether the closeness is significantly different from some norm, will be called the 'outcome-value'.

The third rule is the procedure for minimizing the so called 'explanation effect'. This is the phenomenon of accepting without question the first plausible explanation that is given for a difference between the two values: between the expectation-value ($\hat{Y}_T$) and the book value (Y_B). In the AP situation the 'explanation effect' can arise when the auditor seeks management explanations when significant differences occur for the outcome-rule.

1.7 The Structure of the rest of the thesis.

The thesis is structured as follows. After a consideration of the audit process in Chapter 2, the structure of an AP is addressed in Chapter 3. Chapters 4, 5 and 7 report various findings from the literature review. First, Chapter 4 considers the environment under which APs are used, that is the requirements of the standards and the standard setters. They include the Auditing Practices Board (APB), the Courts and Parliament: the latter, either directly or through agencies such as the Department of Trade and Industry (DTI). The perceptions of those who make use of APs are considered in Chapter 5, which reviews a number of surveys carried out in the USA and the UK that have been published since 1982. That is followed, in Chapter 6, by discussion of the survey conducted for this thesis. The literature review is then completed, in Chapter 7, by an analysis of a number of investigations by academics into the effectiveness of APs. Chapter 8 then reports a number of extensive computer simulations to address identified deficiencies in the earlier experiments and also to provide some new work that looks at the occurrence of false signals when the investigation-rules are used with well controlled data from the pay function of two Government Agencies as referred to in section 1.5. The frequency of false signals is measured for all combinations of the parameters of outcome-value ($\leq 10\%$) and seeded error ($\leq 15\%$). This is done both in terms of frequency of the false signals and under what combination of outcome-rule and seeded error they occur. Finally in Chapter 9, the implications for auditors are considered in respect of their strategies for substantive tests.

Chapter 2: The Role of Audit

2.1 Introduction

One of the goals of the auditor is to provide an opinion either that there are no material anomalies in the financial statements that cannot be explained, or that there is a significant error. To reach either conclusion, the auditor must gather evidence. The audit process can be thought of as an accumulation of knowledge in which different sorts of evidence are brought together to yield an audit opinion. Alternatively, it can be viewed as interactive planning; different sorts of evidence are collected until no more is needed. This process is shown diagrammatically in Figure 2.2.

Auditors are under commercial pressures to minimise their charges and hence their costs. In the light of this the auditor's aim is to collect the minimum amount of evidence that is sufficient to support the audit opinion. Most audits are on a continuing basis and that is useful because, before auditors start the audit, they have a large amount of information about the audited organisation. For example, frequently they will have already completed an audit of the same organisation for the previous financial year. Thus the auditor has a prior opinion based on first hand knowledge. For example, they may suspect that there is sufficient error to mislead the user of the financial statements. In a sense auditors are adopting a Bayesian type argument.

There are many sources of audit evidence, including APs, which are defined in the next chapter. Before discussing the role of APs in the process of evidence gathering, this chapter will outline the audit process and define some key audit concepts. Fundamental to this process are the standards set by auditors for their profession. In the UK these standards are issued by the Audit Practices Board[1] (APB) and are set out in the Statements of Auditing Standards (SASs). The authority of the Standards is set out by the APB as: '*SASs contain basic principles and essential procedures ('Auditing Standards'), which are indicated by bold type and with which auditors are required to comply, except where otherwise stated in the SAS concerned, in the conduct of any audit of financial statements*' (Scope of Statements of Auditing Standards (paragraph 4)).

[1] All the major accountancy bodies in the UK and the Republic of Ireland are members of The Audit Practices Board.

2.2 *What is the purpose of financial statements?*

Financial statements (sometimes called accounts) are prepared for organisations that trade or collect and/or disburse money, ranging from commercial companies and government agencies to private clubs such as sports clubs or those related to hobbies. Such organisations are generally referred to as entities in the Auditing and the Accounting Standards and other professional literature.

European Law, the 4th, 7th and 8th Directives, underpins the statutory nature of financial reporting. The International Auditing Standards are the standards recognised by European Law and the UK Standards comply with those. Each UK SAS has a statement about this compliance. For example SAS 410, on the subject of APs, has the following statement at paragraph 22: '*Compliance with this SAS ensures compliance in all material respects with International Standard on Auditing 520 'Analytical Procedures''*. For the remainder of this thesis we will refer to the UK Standards, recognising that above them is European Law. For example, as with all community law, the 4th Directive has primacy over the UK Companies Acts.

The Audit Practices Board (APB) defines the purpose of the financial statements in the following way: '*Financial statements are normally prepared and presented annually and are directed primarily toward the information needs of an entity's shareholders, proprietors or equivalent body of persons. Generally, financial statements are prepared with the objective that they present a true and fair view of the state of affairs of the entity at the period end and of the profit or loss for that period'* [SAS100]. The persons referred to in this definition might be the shareholders of a publicly quoted company, an individual or group of members at a private company or Parliament for Government Departments or Agencies: the stakeholders. The stakeholders include the wider public. For example there is a statutory requirement for all financial statements to be deposited with Companies House: this forms a database of information for the society at large (e.g. creditors, debtors, potential buyers etc.).

Interestingly, the Auditing Standards do not define a key concept, true and fair, even though there are 317 references to the phrase in the standards and associated guidance. In essence 'true and fair' means that the financial statements present sufficient relevant information that provides the user with all the information they need and that the

financial statements fairly represent the truth. European Community Legislation places a requirement on entities to prepare accounts, which give a: '*true and fair view of the companies assets, liabilities, financial position on profit and loss*'. The Directive does not define this concept.

A very good technical definition of true and fair comes from Hatherly and Skuse (1991). It is: '*Financial statements of an enterprise give a true and fair view if they, together with related notes, are sufficiently informative of matters that affect their use, understanding and interpretation by those for whom they are intended, and they are prepared in accordance with accounting principles appropriate to the circumstances of the business*'. In relation to true and fair, a point that is frequently overlooked is that the primary responsibility for the financial statements belongs to the entity whose affairs they represent. The sequence of events is that the Directors of the entity assert that the financial statements provide a true and fair view of the organisation's financial affairs. The auditor's subsequent responsibility is to provide an independent attestation that the financial statements provide a true and fair view.

In case it might be thought that there is no room for uncertainty or imprecision, the APB make an important point about the non-uniqueness of a set of financial statements. '*A degree of imprecision is inevitable in the preparation of all but the simplest of financial statements because of inherent uncertainties and the need to use judgement in making accounting estimates and selecting appropriate accounting policies. Accordingly, financial statements may be prepared in different ways and yet still present a true and fair view*' (Paragraph 4 of SAS 100).

Although all companies are legally required to prepare financial accounts, not all financial statements are audited. Until 2000, UK law required companies with a turnover greater than £350,000 to have their financial statements audited to attest the truth and fairness of them. From 26 July 2000 companies with a turnover of less than £1m can claim exemption from a statutory audit of their accounts[2].

An example will be used to illustrate some of the concepts discussed in this and subsequent chapters. It is taken from a Government Agency: the UK Passport Agency, which is now the UK Passport Service, whose main source of income is the fees from

[2] Press release from Companies House (Part of the Department of Trade and Industry).

the issue of, and amendments to, passports. There are two items in the income part of the financial statements: the income from the issuing of passports and income from amendments to passports. These are shown in table 2.1. Such items are frequently known as line items, account areas or balances.

Extract from the Financial Statements showing Income	**Income**
Issues of passports	£90,600,000.00
Amendments to passports (e.g. change of name/photographs)	£17,000,000.00
Total Income	£107,600,000.00

Table 2.1: Extract from the Financial Statements relating to Income for the UK Passports Agency for the financial year 1998/1999.

2.3 The Auditing Standards and the Role of Audit

In the UK the Auditing Standards define the objective of an audit of financial statements as follows. *'The objective of an audit of financial statements is to enable auditors to give an opinion on those financial statements taken as a whole and thereby to provide reasonable assurance that the financial statements give a true and fair view (where relevant) and have been prepared in accordance with relevant accounting or other requirements.'* (SAS 100: Objectives and general principles governing an audit of financial statements). They then define how that is to be achieved: *'In undertaking an audit of financial statements auditors should:*

(a) carry out procedures designed to obtain sufficient appropriate audit evidence, in accordance with Auditing Standards contained in SASs, to determine with reasonable confidence whether the financial statements are free of material misstatement;

(b) evaluate the overall presentation of the financial statements, in order to ascertain whether they have been prepared in accordance with relevant legislation and accounting standards; and

(c) issue a report containing a clear expression of their opinion on the financial statements.'

An excellent general definition of the audit process comes from the USA: *'auditing is a systematic process of objectively obtaining and evaluating evidence regarding assertions about economic activity and events to ascertain the degree of correspondence between these assertions and established criteria and communicating*

the results to the interested user' (American Accounting Association (AAA) (1973)). This makes clear the process of obtaining the evidence so that the auditors can issue the report that contains their audit opinion.

The importance of the auditor's work is emphasised by SAS100: *'The auditors' opinion enhances the credibility of the financial statements by providing reasonable assurance from an independent source that they present a true and fair view.'* The APB implicitly recognises that audit cannot provide complete assurance. The Auditing Standards caution *'... the user cannot assume that the auditors' opinion is a guarantee as to the future viability of the entity nor an assurance as to the efficiency or effectiveness with which management has conducted the affairs of the entity.'* (SAS100: Objective and general principles governing an audit of financial statements.)

In a nutshell therefore, when undertaking an audit, the Auditing Standards require the auditor to obtain, sufficient appropriate evidence to enable them, with reasonable confidence, provide an opinion about the financial statements.

2.4 The Audit Process

The process by which the auditor collects sufficient appropriate evidence will be described with the aid of the schematic diagram on the audit process at Figure 2.2 on page 15. By convention there are three phases to an audit: planning, evidence collection and review. These are not distinct: one audit partner during the information gathering phase of this thesis said *'an audit can be thought of as a continuous, evolving, plan where the work changes as a result of the evidence gathered and analysed'.*

The three phases are coloured in the boxes as yellow for planning, green for evidence gathering and blue for the review phase. At any stage in this process, information gathered might suggest that the auditor returns to the planning assumptions and re-assesses the plan and the audit work. The subject of this thesis, APs, is involved at all phases of the audit. But before describing APs in the next chapter we explain some of the concepts involved in the audit process.

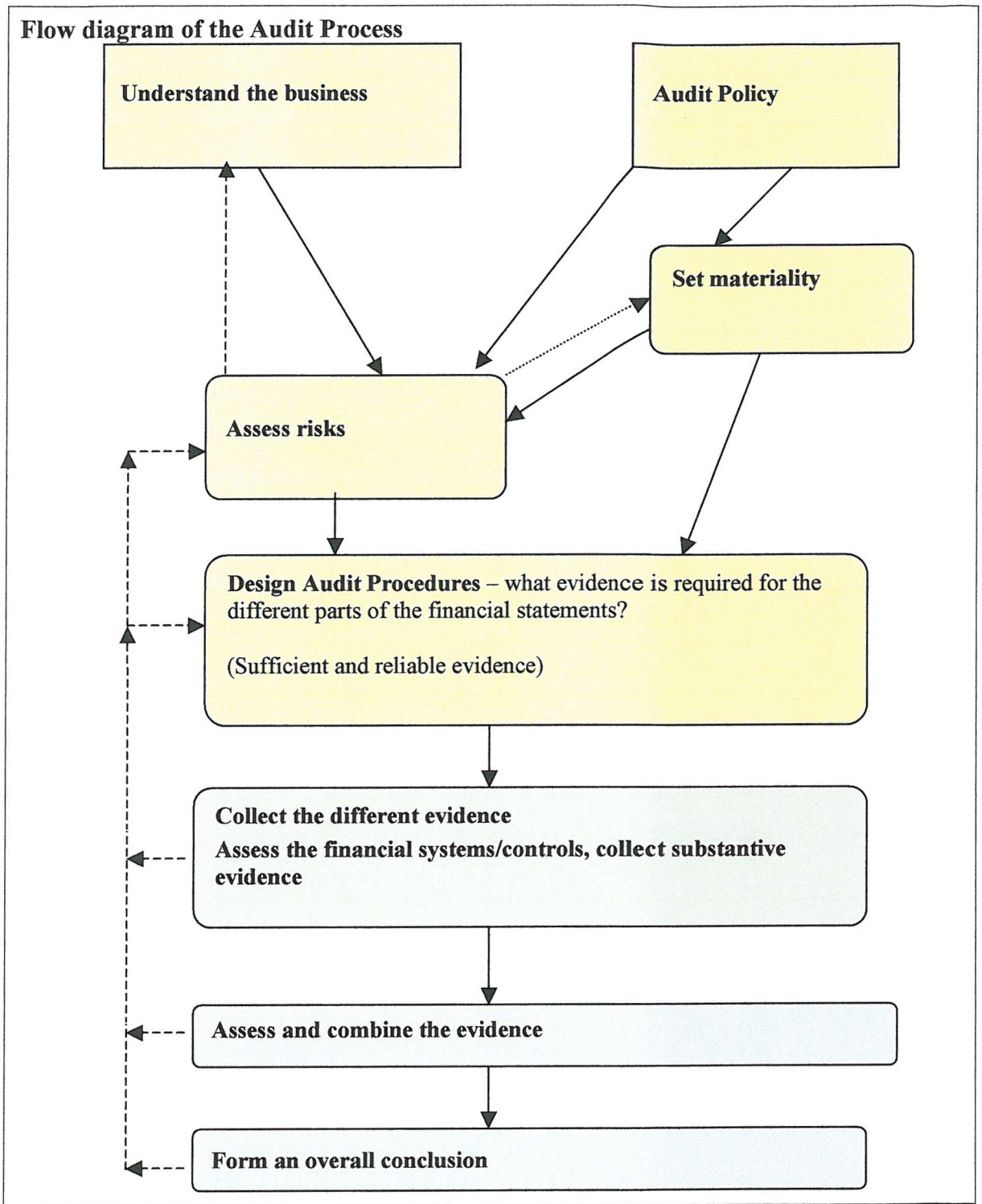


Figure 2.2: Flow diagram of the audit process. At any stage in this process information gathered might suggest that the auditor returns to the planning assumptions and re-assesses the plan and the audit work.

2.4.1 Materiality

Decisions about the nature and amount of audit work are based on judgements about materiality and perceived risks. These are the key concepts that drive the work in an audit. The UK Auditing Standards define materiality to be: '*An expression of the relative significance or importance of a particular matter in the context of financial statements as a whole. A matter is material if its omission or misstatement would reasonably influence the decisions of an addressee of the auditors' report. Materiality may also be considered in the context of any individual primary statement within the financial statements or of individual items included in them. Materiality is not capable of general mathematical definition as it has both qualitative and quantitative aspects*' (Glossary of Terms contained in the Auditing Standards - UK). Clearly auditors assess materiality using their professional judgement. As part of that they should address issues that might influence the decisions of the addressees of the auditors' report. The need for flexibility is also clear from the latter part of this definition. The strength of the auditor's belief that the account figures are within materiality is couched in terms of 'Reasonable Assurance'. For example Glossary of Terms to the Standards contain the statement: '*When reporting on financial statements, auditors provide a level of assurance which is reasonable in that context but, equally, cannot be absolute, that the financial statements taken as a whole are free from material misstatement*'.

As an example the auditor might decide that for the Passport Agency, Table 2.1, the important figure is income. The total income is £107.6 million. The auditor might decide that as long as the income was correctly stated to within £100k (0.1% to one decimal place), then that would not unduly influence the user of the auditors report: Parliament in this case. Another audit might not require such a tight materiality and £Im (0.9%) might be sufficient). However if the auditor knew that the line in the financial statements 'amendments' was of particular interest then the materiality might be based on that figure. It could be set at £0.25 million (1.5% of the line item, to one decimal place). The chosen materialities are arbitrary in the sense they are not based on a formula, but are an expert judgement.

A link to the amount of audit work is contained in SAS 220, which states that: '*Auditors should consider materiality when determining the nature, timing and extent of audit procedures*' (SAS220.2 Materiality and the Audit). SAS 220 goes on to say,

'If auditors identify factors which result in the revision of their preliminary materiality assessment, they consider the implications for their audit approach and may modify the nature, timing and extent of planned audit procedures'. That includes additional work. The Standards go to make the point that a change in the level of materiality may lead to a change in the amount of audit work. *'For example, if, after planning for specific audit procedures, they determine that the acceptable materiality level is lower than was previously assessed, the risk of failing to detect a material misstatement necessarily increases. They compensate for this by carrying out more audit work'* (my underlining). This additional work may be an extension of existing work or entirely different procedures. Thereby the standards are acknowledging the connection between the materiality and the amount of work.

Another key concept, Audit Risk, is defined in the following way: *'Audit risk means the risk that auditors may give an inappropriate audit opinion on financial statements'* (SAS 300). It is the probability that the auditor through their judgements, collection and interpretation of evidence comes to the wrong conclusion. It is not clear from the definition whether this includes the risk of the wrong decision if the auditor concludes there is material error when none exists: the type I error.

2.4.2 Planning the Audit

The audit cycle begins with the planning phase (the yellow boxes in Figure 2.2). Fundamental to the whole process is 'understanding the business'. This informs the important decisions about materiality and risks. Using those assessments to focus on parts of the financial statements that may be at risk of material error or misstatement, the auditor decides what evidence is needed. The amounts and types of evidence may be different for areas perceived to be at risk, as opposed to those not at risk. For example areas not at risk may only need some simple work, while an area at risk may require the testing of all transactions of a particular type or sampling.

In the Passport Agency example, the auditor might decide that although the income from amendments to passports is not the largest source of income (at £17 million it is less than 16% of the total income), it is particularly at risk. There may be many reasons for that, for example poor controls. For that reason the risk of error in 'Income from Amendments' might be set at high, and it might be given particular attention.

2.4.3 Evidence gathering: auditing the systems of control

At the evidence-gathering stage – the green box in figure 2.2 – there are two strategies that might be employed. They are an assessment of the reliability of management's systems of control and/or substantive procedures. If the auditor decides to place reliance on the systems, then as well as reviewing those systems to assess whether they are working properly, the auditor will identify and test key controls.

2.4.4 Evidence gathering: Substantive Procedures

Another part of the evidence-gathering phase is Substantive Procedures. These are described in the Glossary of Terms in the Auditing Standards as – '*tests to obtain audit evidence to detect material misstatements in the financial statements. They are generally of two types: firstly APs; and secondly other substantive procedures, such as tests of details of transactions and balances, review of minutes of directors' meetings and enquiry*'. (Statements of Auditing Standards (1995)). Such tests are always required, as the Standards make clear: '*The nature, timing and extent of substantive procedures depends, amongst other factors, on the auditors' assessments of the control environment and accounting systems generally and of the inherent and control risks relating to each assertion, as well as on any evidence obtained from audit work performed during the preparation of the financial statements. In particular, where tests of control provide satisfactory evidence as to the effectiveness of accounting and internal control systems, the extent of relevant substantive procedures may be reduced, but not entirely eliminated*'. (SAS400 Audit Evidence, paragraph 14). The SAS requires that there should be sufficient and appropriate evidence. What constitutes sufficient and appropriate evidence is described in the following way. '*Sufficiency and appropriateness are interrelated and apply to audit evidence obtained from both tests of control and substantive procedures. Sufficiency is the measure of the quantity of audit evidence; appropriateness is the measure of the quality or reliability of audit evidence and its relevance to a particular assertion. Usually, audit evidence is persuasive rather than conclusive, and auditors therefore often seek audit evidence from different sources or of a different nature to support the same assertion. Auditors seek to provide reasonable, not absolute, assurance that the financial statements are free from material misstatement. In forming their audit opinion, therefore, auditors do not normally examine all of the information available. Appropriate conclusions can be*

reached about a financial statement assertion using a variety of means of obtaining evidence, including sampling'. (SAS 400: Audit Evidence). This makes clear that total assurance is not a realistic goal.

We can illustrate how understanding the business might drive the audit. Information gathered as a result of reviewing the systems should improve the auditor's understanding of the business. That additional information might change the understanding of the business and that in itself could lead to a change of plan. That illustrates the iterative nature of the audit process – in doing the audit, the auditor should gain a better understanding of the business, to plan the audit an appreciation of the business is also necessary. The auditor obtains more evidence and information as the audit develops. Thus their knowledge of the business develops and that should feedback into audit work: changing its nature and improving the quality of the audit.

In the Passports Agency example, if the systems controlling the 'Issue of Passports' were considered to be good, then it might be that the audit would focus on those systems and there would be a reduction in the amount of substantive work for that line of the financial statements. The systems controlling the 'Amendments to Passports' account area might not be so strong or there might be a perceived risk, so it would be the subject of some focussed substantive procedures. It might be some form of stratified sampling or a combination of APs and sampling. So the two areas of the income part of the financial statements might have different audit approaches.

2.4.5 Review of the audit

The final phase of the audit is to bring the audit evidence together and to assess whether the different parts of the financial statements fit together and are consistent. The Standards state that: *'Auditors should carry out such a review of the financial statements as is sufficient, in conjunction with the conclusions drawn from the other audit evidence obtained, to give them a reasonable basis for their opinion on the financial statements'*. (SAS 470.1)

2.5 Combining Audit Evidence: A Short History of the Audit Risk Model

2.5.1 Introduction.

The purpose of this section is to put into context the rationale auditors employ to combine their audit evidence and bring together all the signals from the various pieces of audit work. The rationale can be characterised as a framework, or structure, for combining the information/data/evidence from the different stages of audit work. A common structure for this purpose is known as the Audit Risk Model (ARM). There are references in the literature to the Audit Risk Model dating back to the early 1960s, although there is evidence of its application[3] in the 1950's. This section traces the history, and suggests why the terminology and concepts have confused the issue of combining evidence. It is discussed in more depth than some of the other concepts because some of the work described in Chapter 7 depended on the Audit Risk Model.

2.5.2 The auditor's decomposition of the risks

The Auditing Standards provide the auditor with some guidance on the assessment of risks and on combining the evidence from different sources. The relevant risks are defined in a glossary of terms attached to the Auditing Standards. The abbreviations AR, IR, CR and DR will be used in the rest of this chapter.

- **Audit Risk (AR)** – The risk that auditors may give an inappropriate audit opinion on financial statements. Audit risk has three components: inherent risk, control risk and detection risk.

- **Inherent Risk (IR)** – The susceptibility of an account balance or class of transactions to material misstatement either individually or when aggregated with misstatements in other balances or classes irrespective of related internal controls.

- **Control Risk (CR)** – The risk that a misstatement that could occur in an account balance or class of transactions and that could be material either individually or when aggregated with misstatements in other balances or classes would not be

[3] As part of a survey of Audit Firms I visited the New York Headquarters of Coopers and Lybrand (USA). In a discussion with A. J. Lorie, a partner of Coopers and Lybrand, he described the use of a simple form of the Audit Risk Model.

prevented or detected and corrected on a timely basis by the accounting and internal control systems.

- **Detection Risk (DR)** – The risk that auditors' substantive procedures do not detect a misstatement that exists in an account balance or class of transactions that could be material either individually or when aggregated with misstatements in other balances or classes.

The type of substantive procedure used by the auditor and the amount of work are driven by their assessment of inherent and control risks. The SAS 300 on Accounting and internal control systems and audit risk assessments says that: *'Auditors should consider the assessed levels of inherent and control risk in determining the nature, timing and extent of substantive procedures required to reduce audit risk to an acceptable level'* (SAS 300.7). The International Standards make the clear connection between the assessment of risk and extent of the substantive procedures: '*the higher the assessment of inherent and control risk, the more evidence the auditor should obtain from the performance of substantive procedures*' (International Standard on Auditing 400 'Risk Assessments and Internal Control').

2.5.3 Purpose of an Audit Risk Model

It is clear that some structure/rationale can be helpful in bringing together the results of different audit tests and other information that the auditor acquires during the progress of an audit. Research indicates that most humans have difficulty in integrating the effect of several interrelated factors that are relevant to their decision about, in this context, the truth and fairness of a set of Financial Statements. For example, Libby commented on *'the limited ability of people to integrate information from different sources appears to be the most consistent finding of the literature reviewed in this [his] book'* (Libby (1981)). One of the pieces of research that Libby reviewed[4] came to the conclusion that '*experts are much better at selecting and coding information than they are at integrating it'*.

There is a need to ensure that auditors make consistent judgements when planning an audit and making a decision on the acceptability of an audit result. The auditor has

[4] R M Dawes [1979]: The Robust Beauty of Improper Linear Models in Decision Making. American Psychologist (July 1979).

many different sources of evidence, including judgements they make about the organisation, the audit work on the adequacy of the management's systems of controls and the results of tests of detail. There is a need to ensure that evidence is combined in a consistent and defensible way: from audit to audit, and between auditors on a particular audit.

The Audit Risk Model (or an Audit Assurance Model) is one attempt to achieve consistency. It's validity is disputed by some, but O'Hagan summed up the requirement for some process: '*Of course the auditor's knowledge and experience, and the results of previous audit activities, should not be wasted when it comes to determining the substantive sample size*' (van Batenburg, O'Hagan and Veenstra (1994)). Aldersley (1989) made the point that even if the auditor fails to recognise the practical limitations of the risk model as an operational formula, they will at least be complying in some qualitative sense with the spirit of the requirement to consider overall risk. Steele (1992) made a similar point in commenting on the necessity for some form of model '*although the risk model does need some special assumptions in order to be logically coherent, it does try to communicate an important truth about the combination of judgement and evidence*'. So in the light of these comments what is the Audit Risk Model and what are its origins?

2.5.4 Intuitive Use of the Model for combining information

Before describing the models, it is worthwhile considering the auditor's use of the word 'Risk'. This appears to have two different meanings within the audit context. It could be the risk attached to a part of an account or organisation. For example debtors might pose more risk than payroll. Within a firm a particular branch might pose more risk than the others. This is the organisational risk or Entity Risk. The other meaning of risk refers to the risk that the auditor's own procedures will fail to detect/identify significant error (should it exist). This is the auditor's personal risk or Auditor Risk. This distinction is important since the Audit Risk Model contains variables, which are different aspects of risk as it affects the organisation being audited on one hand and the work of the auditor on the other.

2.5.5 Evolution of the Current Risk Models

The current models developed over a period of time. They started out as extremely simple models and were refined in an attempt to model the complexity of the audit process. The significant developments of the Audit Risk Model are described below. They represent attempts to take account of the different stages/aspects of the audit process. Earlier models based on just two components, were decomposed in attempts to reflect more realistically the different parts of the audit process.

Mautz and Sharaf (1961) published a monograph on the 'Philosophy of Auditing', which reflects the simpler approach to auditing of the 1950's. It provides an excellent perspective of the rationale of auditors in that era. For example it is clear that much of the audit thinking focused on the application, and extent, of sampling to test controls and sampling for the tests of detail. Perhaps because of this concentration on sampling, the concepts of confidence and probability were the focus of much of the work. Mautz and Sharaf considered the audit opinion, at an overall level, and at an account area level, in terms of 'probability'. When discussing whether the *'customary tests and samples on which the auditor relies are sufficient to justify the audit opinion'*, they pointed out that *'it must be recognised that we are dealing with probabilities'*. Here they were thinking about the overall probability that the audit work, of which sampling may be a part, is sufficient for the auditor to conclude that the accounts are free of material error or that there is material error. The aggregate of the work, if it is not contradictory, tends to increase the probability that the proposition can be accepted with little chance of making a false decision: *'the combination of types of evidence all in support of the same proposition'* (Mautz and Sharaf (1961)).

From this it was natural for auditors to think in terms of a simple model that dealt in probabilities: those associated with the sampling at the controls testing and substantive testing phases of an audit. The simplest model was: $AR = CR \times DR$, where $0 \leq AR \leq 1$. The intention was that AR would be close to 0, for example ≤ 0.05. This can be considered as a model of two important parts of the audit work: the identification and testing of the systems of control and sample based substantive tests. It was simplified model of the whole audit process. The model was first promulgated by the AICPA in 1972.

Carmichael (1968) also focused on sampling as a specific test, and consequently the discussion of audit tests was in terms of 'confidence' achieved. In some published works, the reliability of a procedure is treated as if it were equivalent to the confidence associated with sampling and the concepts Risk / Confidence / Probability became interchanged with each other. It was not until the 1970s that the focus started to shift from sampling (and confidence) to audit risk as a broader concept covering, in a qualitative sense, the likelihood of the elements IR and CR permitting the occurrence of undetected material error. In practice, a further difficulty for the current model is that the different components are not really probabilities but are numbers assigned to the different sources of information: controls assessment, inherent risk assessment, APs and other substantive tests. They are set in a way that, for auditors, produces logical combinations and weightings of the different parts.

In 1975 Stringer, who had pioneered the use of regression as an AP, proposed an enhanced model that split the assurance arising from substantive work into the AP and sampling elements and so disaggregated the Detection Risk into a risk that the APs would not detect material error (Ar) and a risk that the sampling would not detect material error (SR). That was a logical development, attempting to reflect the different types of work being done: AR = CR × Ar × SR. Some researchers hypothesised that the Control Risk element included Inherent Risk, the propensity of the organisation to allow material error. Anderson (1977) proposed a model that had inherent risk (IR) as a component. The model became: AR = IR × CR × Ar × SR.

It is at this point, when IR was separated out from control risk, that the model ceased to be a barometer of the auditor's own work and attempted to encapsulate the whole audit process; from the propensity of the accounts to contain significant error, to the auditor's chance of detecting such events. Thus the model had started to reflect the Entity (organisation) being audited and the auditor's own work on the controls and the substantive tests. However, it still did not reflect all the auditor's different sources of evidence.

Leslie, Teitlebaum, Anderson (1979) proposed a model which allowed the auditor to input judgements of the likelihood of 'material' significant error occurring within an organisation. The model was AR = CR × SR × ME, where ME represents the likelihood of material error, subjectively assessed by the auditor. In a way this concept

of ME is similar to IR, it is making the auditor focus on the likelihood of error rather than the less specific definition of inherent risk. This appears to have gained no acceptance.

In 1981 the American Institute of Certified Public Accountants (AICPA) published, in SAS 39, a model that reflected Stringer's ideas. It was AR = CR × Ar × SR. Two years later the model changed to AR = IR × CR × DR (AICPA in SAS47: 1983). Comparing the Anderson model to the 1981 AICPA model or the Stringer Model (1975) we can see that they are all refinements of a common model. This can be seen if IR is set to 1, then the Anderson model is the same as 1981 AICPA model or the Stringer Model (1975). Table 2.3 below shows all these models for convenience.

Originator of Model	Model
AICPA (1972)	AR = CR × DR
Stringer (1975).	AR = CR × Ar × SR
Anderson (1977).	AR = IR × CR × Ar × SR
Leslie, Teitlebaum, Anderson (1979).	AR = CR × SR × ME
AICPA (1981) - SAS39.	AR = CR × Ar × SR
AICPA (1983) - SAS47.	AR = IR × CR × DR

Table 2.3: This shows the different Audit Risk Models discussed in this Chapter.

These are just a few of the models that have been developed over the years: they all evolved from one simple model.

2.5.6 Necessity for some form of Audit Risk Model and some criticisms

In attempting to combine different evidence sources in a consistent and coherent way, a structure, framework or model is useful. Many researchers felt that, although the audit risk models were deficient, they did offer a mechanism that provides some structure to assist with the combination of evidence. For example, in relation to the model AR = CR × DR, Andrews and Smith (1987) said: '*Although the events to which these risks apply are not clearly defined the ensuing rules are sensible and lead to the decision that small samples will suffice when the system is good but large samples are needed when it is bad*'. Andrews and Smith went on to construct a new model that explicitly took account of the prior information the auditor had. At each stage there would be an information gain: from planning work, assessing the systems of control, substantive

testing etc... Also at each stage the model took account of the information already gathered and incorporated the new information. It was a Bayesian formulation and attempted to assess how much more information was required at each stage to ensure that the auditor's assessment of the level of potential error in the financial statements was less than some critical value such as materiality.

From the early stages of the development of the Audit Risk Model doubts were raised about whether such a model represented the way the audit process functions. As an early example of such concern, Smith (1972) commented that: *'No logical basis has been determined for setting the confidence level correlated with the different states of internal control. The selection of levels to be utilised is completely arbitrary'* (Smith (1972)). His 'internal controls' are the same as the 'accounting and internal control systems' in the definition of control risk. Although Smith's comments were in respect of measuring the contribution of the assessment of the systems of control (CR), relative to the rest of the audit work, the comments are also valid about the IR and Ar components of the Audit Risk Model. In those cases there is no mathematical measurement of the risks, as there might be from a statistical sample where we can assess the probability of the result being from a sample that is not representative of the population being audited.

Some authors have taken the view that the Audit Risk Model never reflected the way auditors combined information. This can be illustrated by looking at the work of Jiambalvo and Waller (1984). They carried out an experiment to investigate whether the auditor might make better judgements about setting the amount of substantive testing that should be undertaken to audit a particular value within a set of financial statements, if overall audit risk (AR) was decomposed into the component risks associated with controls work (CR), and with the analytical review work part of substantive testing (Ar). The auditor then decided on AR, CR and Ar before assessing test of detail (SR) as part of substantive testing.

The experiment involved 13 qualified auditors each with at least four years audit experience. They were from one of the major audit firms in the USA. The auditors were split into two groups, both using the same data, contained in four case studies. The experiment required the first group (6 auditors) using the information in the case studies to consider the components of the Audit Risk Model individually and then

assess a risk for each of AR, CR, Ar, before making an assessment of SR in each of the four case studies. The second group of seven auditors were required to make a holistic assessment of SR in each of the four case studies. Jiambalvo and Waller then compared the assessments of SR of the two groups and concluded that there was not a statistically significant difference between the two groups assessment of SR. This implies that the auditors did not use the information gained from the decomposition to help decide SR. This may be because they did not view the Audit Risk Model as a means of determining SR, or it could be because the experiment was ill founded.

Jiambalvo and Waller then carried out a second analysis using the data from the first group. The comparison was of the auditor's intuitive assessment of SR to the value of SR calculated by using the model AR = CR × Ar × SR (i.e. SR = AR / (CR × Ar) utilising that group's intuitive assessments of AR, CR and Ar. There were significant differences. The calculated values for SR were, with 2 exceptions, greater than 1 and for many cases much greater than one: in 7 out of 24 case studies they were ≥ 10. Clearly, the auditor's assessments were not compatible with the concepts behind the Audit Risk Model. The results of the experiment, however limited it was, were particularly worrying in that, at the time of the experiment, the audit firms made extensive use of the Audit Risk Model. In the light of that, Jiambalvo and Waller 's results therefore suggest that many auditors do not understand the concepts behind the methods they use. The conclusion of the paper was that: '*assuming the absence of error in eliciting the auditors assessments of UR, IC, AR, and TD risks, it is clear that the auditor's intuitive combination of the risk components did not correspond with that predicted by the Audit Risk Model*' (Jiambalvo and Waller (1984)). The UR, IC, AR, and TD risks were respectively audit risk (AR), control risk (CR), analytical risk (Ar) and sampling risk (SR) as defined on page 20.

The result that decomposition did not appear to improve the auditor's judgement appears, at first sight, to contradict Libby's (1981) argument that such a decomposition process would aid the auditor's judgement. But if we consider the hypothesis that the Audit Risk Model does not replicate the audit process and thus may not reflect the way auditors combine information, we see that Jiambalvo and Waller's results are not inconsistent with Libby's ideas. Daniel (1988) lends support to this hypothesis. Her analysis of an experiment indicated, *'the auditors did not combine the components of*

audit risk in a manner suggested by any of the authoritative models'. She also observed that it is '*possible that there are other factors or components in audit risk decision that have not been properly identified and addressed in existing models'* (Daniel (1988)).

The conclusion that the Audit Risk Model does not reflect the auditor's process of combining information and making judgements forms part of the argument of O'Hagan and van Batenburg (1994) who described the Audit Risk Model in terms of Assurance calling it an Audit Assurance Model, assurance being the converse of risk. They put forward the proposition that: *'the Audit Assurance Model is a statistically doubtful model, containing variables that should not be in it, with numerical values that cannot be validated, and giving results that are methodologically invalid. Of course the auditor's knowledge and experience, and the results of previous audit activities, should not be wasted when it comes to determining the substantive sample size. Some variables in the Audit Assurance Model are good ways to describe the auditor's 'professional judgement'. However those variables do not affect the confidence level used to test for the presence of error, but should influence the distribution of the level of error itself'* (van Batenburg, O'Hagan and Veenstra (1994)). This is the formation of the model that Andrews and Smith (1987) addressed in their paper, which was referred to on page 25.

All of these criticisms have some substance. The Audit Risk Model has developed significantly from its origins of over forty years ago when it was about the combining the results of two sampling exercises: one to test the systems of control and one to carry out a substantive test. It was not a model of the audit process. Over the years auditors have tried to make it so by decomposing the controls risk into IR and CR and the risks from different types of substantive test into Ar and SR. The recent models no longer reflected the original simple concept. But in addition to these developments being, there is an issue that few of the models attempted to address. They assume that the relative worth of the different audit evidence is of similar value. This leads to the question as to whether control evidence is stronger than substantive evidence, or vice versa?

Andrews and Smith (1987) were trying to broaden the argument: when forming the overall conclusion, in addition to the amount of assurance achieved, they considered another attribute: how strong should the evidence be? Such a strategy would be consistent with the ideas of Mautz and Sharaf (1961), who put forward the proposition

that more compelling evidence is required for material assertions than for those that are not material. They drew a parallel with US law, which distinguishes between two degrees of proof. For civil cases those propositions must be established with probability of over one half, or on the basis of *the preponderance of evidence.* For criminal cases, on the other hand, the truth of a proposition requires a degree of probability that differs from certainty by so little, that anyone acting on that difference would be regarded as unreasonable; such strength of evidence is termed *proof beyond reasonable doubt.* Mautz and Sharaf proposed that the preponderance of evidence criterion should apply to assertions that are not material, while the proof beyond all material doubt is required for material assertions. It follows that a single source of evidence for a material assertion must be compelling or a combination of different types of evidence must be obtained. All the audit evidence should indicate a high probability of the truth of any material assertion.

2.5.7 Reliability Theory Approach

If we consider reliability theory, we can see some parallels with the structure of the Audit Risk Model. The theory of reliability conceives that an output as being built up from components in a sequence. At each stage the outcome is a zero / one component. An example is in the Figure 2.4 below.

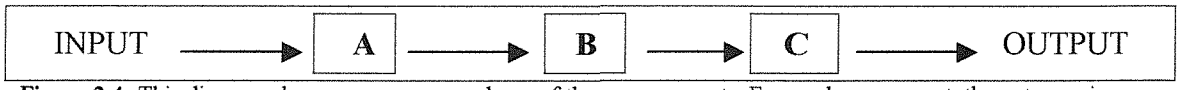


Figure 2.4: This diagram shows a sequence made up of three components. For each component, the outcome is either zero or 1.

If a component fails, the system fails. Each component can fail with probabilities p_A, p_B, and p_C, all which are assumed to be independent. We can write the combined probabilities in the following way.

$$\begin{aligned} \text{Pr (Output Achieved)} &= \text{Pr (no component fails)} \\ &= (1 - p_A)(1 - p_B)(1 - p_C) \\ \text{Pr (System Fails)} &= 1 - (1 - p_A)(1 - p_B)(1 - p_C) \end{aligned}$$

A question is whether this represents the audit process? The form of the Audit Risk Model is similar: AR = IR × CR × DR. But this format is not the way the audit process works. In reliability theory if any component fails, the system fails. With the audit process, if the audit work on the systems indicates that they cannot be relied upon then

the result is not, necessarily, that the audit has failed. Rather alternative evidence is collected to confirm or deny the evidence from the systems audit. Thus the audit process does not follow reliability theory although the ARM appears to use that rationale.

The audit process is an accumulation of information from a variety of sources; it is not a reliability type system. What is required is some process that allows the combination of many sources of evidence, both qualitative and quantitative, is a structured way. One such approach could use Bayes Theory. In principle such methods allow the auditor to use their prior information from previous audits, other audit work they have done and their knowledge of the organisation to influence how the substantive phase of the audit is conducted, and to infer an audit conclusion from such work and the prior information.

2.6 Conclusion

There are three criteria the auditor must meet in order to provide assurance to the users of a set of financial statements about the 'truth and fairness' of those financial statements. The first is that the financial statements are presented clearly and are in accordance with the relevant legislation. The second is that the auditor obtains with reasonable assurance, sufficient reliable evidence that the financial statements are free of material error. And the third is that the opinion of the auditor is unambiguous. This thesis is concerned primarily with the second requirement, and in particular with one source of evidence – APs and whether APs can provide sufficient, reliable evidence.

Section 2.5 described how many auditors have tried to combine audit evidence. While there is not yet a universally agreed methodology, some strategy is required for demonstrating that the combination of the different sources of evidence is sufficient to satisfy the 'proof beyond all reasonable doubt' criterion. We demonstrated that the Audit Risk Model is an imperfect way of combining the audit evidence and conclusions from its application may be flawed. This will be relevant to some of our considerations in Chapter 7.

In order to set the scene for the rest of the thesis, the next chapter will provide a description of APs and the different categories of AP. It will also propose a formal structure for APs.

Chapter 3: A Description of APs and their Role within the Audit Process

3.1 Introduction:

This chapter describes the basic concepts of an AP. As part of the analysis, APs are categorised into four broad groups, since it is more effective to compare a limited number of subsets, rather than many individual procedures. The categorisations are chosen to be consistent with earlier pieces of research, thereby making it easier to compare the findings with those studies.

3.2 What are APs?

This section discusses the concepts underlying APs, first by giving examples of typical APs, then by defining and discussing APs. In broad terms, APs involve thinking about relationships between the figures within, and data from, the financial statements being audited. APs may also use data external to the financial statements, or even to the entity. In using an AP the auditor should be addressing the question: 'Is this figure credible in relation to these other pieces of information?' rather than asking 'Is this figure credible?' APs can be conceptualised as a framework for data extraction, analysis and description which provides information to the auditor thereby allowing him/her to make judgements about a set of financial statements, or an item within those statements. The methods include comparison and predictive analysis as well as descriptive summaries.

An AP can be thought of as an indicator function that provides a signal that there is not, or there is, material error. The function takes the value '0' if the AP indicates that there is no evidence of material error, and '1' if there is evidence of material error. This concept will form the basis of the analysis, of the effectiveness of APs in providing reliable evidence.

3.2.1 Examples

A few examples are set out below that may give an indication of the ideas that underlie APs, and the variety of available techniques.

- The income from fees for licences should be roughly in proportion to the number of licences issued, perhaps with an adjustment for the composition of the mixture of licences of different values.
- An organisation's pay costs in a given year should be similar to those in the preceding year, modified to take account of wage inflation and of changes in the numbers and grades of those on the payroll.
- The expenditure at a particular location within an organisation should be roughly proportional to some measure of the volume of activity at that location.
- The income from the sale of passports can be predicted by using information about published fee rates and data from the controls of the issue of blank passport books
- The amount paid out in Child Benefit should be proportional to the number of eligible children.

Taking the last example, the need to understand the business is crucial. It is not just the number of children that is important, but also obtaining data on who is eligible. For example, it will include children of those overseas servicemen who fulfil certain requirements.

First we extend the example given at the start of Chapter 2. It will be used to illustrate some of the concepts discussed in this chapter in a situation favourable to the use of APs. The UK Passport Agency's main source of income is the fees from the issue of, and amendments to, passports. The fees for the three types of passport and the passport amendments are set by Parliament; they are shown in Table 3.1. The auditor wishes to predict the income for the financial year. There are two items in the financial statements: the income from the issuing of passports and income from amendments to passports. These are shown in table 3.2.

Passport type	**Fee**
Standard adult 10 year passport (32 pages)	£28.00
Large size adult 10 year passport (48 pages)	£38.00
Child's passport (0 - 15 years, valid for 5 years)	£14.80
Amendments (e.g. change of name/photographs)	£17.00

Table 3.1: The Passport fees authorised by Parliament. [Source: Passports Agency Web Site – May 2001].

Extract from the Financial Statements	Income
Issues of passports	£90,600,000.00
Amendments to passports (e.g. change of name/photographs)	£17,000,000.00
Total Income	£107,600,000.00

Table 3.2: An extract from the Financial Statements relating to Income to the Passports Agency. [Source: Passports Agency Web Site – May 2001].

3.2.2 Some definitions

The two types of substantive procedure are tests of detail and APs. The first deals with individual transactions, whereas APs concern data that have some degree of aggregation. Knechel (1988a) encapsulates the rationale of APs with the statement: APs are '*a substantive auditing procedure that examines the accuracy of accounting balances without considering the details of individual transactions which make up the accounting balance*'. This clearly makes the point that APs deal with aggregated data. Another feature of APs is the formation of an expectation value based on the relationships between different systems and their data. According to Knechel (1988a): *'Analytical review procedures rely upon an auditor's ability to generate an expectation of an account based upon knowledge of a company's business and operating environment.'*

In the same year, 1988, a statement by the AICPA, set out clearly the requirements for a good AP: '*APs involve comparisons of recorded amounts, or ratios developed from recorded amounts, to expectations developed by the auditor. The auditor develops such expectations by identifying and using plausible relationships that are reasonably expected to exist based on the auditor's understanding of the client and the industry in which the client operates*' (AIPCA - SAS No. 56 (1988)). The two key words are comparison and expectation. The former means the estimation or measurement of some similarity or dissimilarity between two figures. The latter is a predicted value of a relevant variable. A very simple AP may involve the comparison between two values. A potentially more powerful AP may involve the development of an expectation and then a comparison. Both involve comparison but the latter involves the explicit development of an expectation value.

The Auditing Standard for the UK and Ireland is less explicit. The Auditing Standard on APs sets out the basic description of an AP, which is: *'The analysis of relationships: (a) between items of financial data, or between items of financial and non-financial*

data, deriving from the same period; (b) between comparable financial information deriving from different periods or different entities; or (c) to identify consistencies and predicted patterns or significant fluctuations and unexpected relationships, and the results of investigations thereof' (SAS 410).

As in the USA Standards, the UK guidance has the concepts of comparison and prediction. However, the USA Standards set out the structure for the expectation process and the need for a framework to ensure that the auditor establishes an estimate of the expectation values in a structured way. The UK Standards tell us what an AP is, but do not say how it will work. Clearly the effectiveness and reliability of APs will depend on the quality of the methods and the data used. These are within the control of the auditor. The UK Standards describe APs as the analysis of relationships. Given sets of data, and relationships that are unique to a particular entity, the performance of an AP by one auditor should produce a similar (comparable) result to that of another auditor. The application of an AP and its outcome should not be arbitrary. That requires having a framework for developing an expectation value.

3.2.3 The two essential features of an AP: Expectation and Precision

The definitions in 3.2.2 highlight the need for the auditor to form an expectation value as a critical stage of the development of the AP. They also imply some concept of measurement of the acceptability of the auditor's estimate, or expectation value. Expectation can be taken to be exactly what it says; the value the auditor expects to be in the audited statements based on knowledge of the business and the relationships between the different data sets. For this thesis, the expectation process is defined to be a statement of the auditor's understanding of the relationship between the explanatory variables and the response variable (the account figure being audited). The result of that formulation is a value or range of values for the account figure. It may be an *explicit* formulation, although frequently the assumptions behind the formulation are only *implicit*.

The examples given in 3.2.1 all imply that the account figure being audited should, in some sense, be 'close' to what is 'expected' when other available information is taken into account. That is, the acceptability of the expectation value should include some measurement of its accuracy: for example, how precise is the expectation value. For

statistical models there are direct ways of measuring the precision. If informal models are used, then the estimate of precision must be based on the auditor's judgement. In practice, auditors use informal methods and exercise judgement in determining what figure they would expect, or what they think is reasonable. The essential difference is that with a statistical model the auditor has a quantifiable value on which to base his / her judgement. It may be that in some situations (such as the use of APs in audit planning) there is no need for a formal definition of 'close'. But if the auditor is to base an opinion on the evidence produced by the substantive procedure, some measurement of closeness of the expectation value to the figure being audited is required. That is, some measure of what is a 'significant difference' is required. In this thesis we use 'significant difference' to refer to one that is important for some reason. We distinguish that from a 'statistically significant difference' where the probability is very low of getting a result (a value) that is as extreme as the one obtained. Both are important and these issues will be addressed later in the thesis in Chapter 8.

The UK Standards do recognise that the accuracy is a factor in the use of AP. *'APs can also help provide evidence regarding completeness. In particular, predictive APs can often be an effective means of testing for completeness, provided the results can be predicted with a reasonable degree of precision and confidence. Variations from expected results may indicate possible omissions which have not been detected by other substantive tests'* (Practice Note 13: 'Audit of Small Businesses', September 1997). The introduction to PN13 explains that: *'The purpose of Practice Notes issued by the Auditing Practices Board is to assist auditors in applying Auditing Standards of general application to particular circumstances and industries. They are persuasive rather than prescriptive. However, they are indicative of good practice, even though they may be developed without the full process of consultation and exposure used for Statements of Auditing Standards'.*

The Auditing Standard on APs (SAS 410) recognises that the extent of reliance auditors place on the results of substantive APs may depend on *'the accuracy with which the expected results of APs can be predicted'.* (SAS 410, paragraph 15(b)). The Standards then go on to say, *'However, reliance on the results of APs depends on the auditors' assessment of the risk that the APs may identify relationships as expected whereas, in fact, a material misstatement exists'.* (SAS 410: APs). The Auditing Standards

therefore recognise that precision and reliability of the AP are issues, but offer no further guidance than their impact is a matter for the auditor's judgement.

3.3 When are APs used within an audit?

A flow diagram of the audit process was given in Chapter 2. In this section of Chapter 3 we discuss APs in the context of the audit process, which can be thought of as an iterative process that continues until there are no identified anomalies to explain. Since the evidence requirements are not the same at each phase of the audit, it is convenient to follow the auditor's classification of the process into three phases: planning, implementing the planned work and review.

At the initial planning and review phases, the use of APs is mandatory. The Auditing Standards require that '*Auditors should apply APs at the planning and overall review stages of the audit*' (SAS 410.1). The purpose of APs is to direct attention to issues that the auditor should investigate or at least consider because of the likelihood of potential problems. The use of APs at this stage of the audit also aids understanding of the entity's business, and assists in planning the audit investigations. '*Application of APs may indicate aspects of the entity's business of which the auditors were previously unaware and assist in determining the nature, timing and extent of other audit procedures*'. *(SAS 410, paragraph 11).* APs at this stage are frequently based on interim financial information, budgetary information and management accounts, using high level data (data at an entity level). For example Loebbecke and Steinbart (1987) reported the use of very simple techniques using annual, not disaggregated, data.

Following the initial planning work, the auditor usually examines the systems producing the financial statements and other business information pertinent to those statements. As a result of this examination and the initial planning work, a decision is made regarding what other evidence is required. Auditors call the determination and gathering of this evidence 'substantive procedures': it is a key phase of an audit. The purpose is to obtain audit evidence to detect whether there are material misstatements in the financial statements. There are two categories of substantive procedure, APs, and other procedures, such as tests of details (requiring inspection of documents) of transactions and balances. APs often provide sufficient substantive audit evidence on their own, but may also be used to supplement other substantive procedures. A typical

use of substantive APs is to test for completeness of income provided the results could be predicted with a reasonable degree of precision and confidence. Variations from expected results might indicate possible omissions that have not been detected by other substantive tests.

The final step is overall review, where APs confirm that there is no issue remaining. The conclusions drawn from the application of such procedures should normally corroborate conclusions formed during the audit regarding individual elements of the financial statements. Of course they may also identify areas requiring further work. The Auditing Standards describe the role of APs at the review stage as producing *' results which assist in arriving at the overall conclusion as to whether the financial statements as a whole are consistent with their knowledge of the entity's business'* (SAS 470.3). The Auditing Standards recognise the importance of APs at this stage of the audit: - *'When completing the audit, auditors should apply APs in forming an overall conclusion as to whether the financial statements as a whole are consistent with their knowledge of the entity's business'* (SAS 410.3).

The auditor is encouraged to use simple procedures at the planning stage to help their understanding of the business. Later in the audit cycle, APs are developed with the aid of the auditor's understanding of the business in order to gather evidence. This apparent contradiction is no more than apparent. The roles of the AP at both planning and substantive stages involve using the relationships within the data to explore when planning the audit and to predict at the substantive stage. These roles are separate but complementary.

There is an analogy with the analysis of data by a statistician. The first step is data exploration, to help the statistician's understanding of the source of the data. That understanding assists the statistician in deciding upon an appropriate method of analysis. The analysis is then carried out. A range of statistical procedures may be used, ranging from simple descriptive statistics such as calculating means and variances to possibly fitting a complex statistical model, if that is appropriate. Different statistical methods at each stage: in general simple for the initial analysis to understand the data and the problem, more complex methods if appropriate to solve the problem. Chatfield (1985) described such an investigation strategy as Initial Data Analysis (IDA).

3.4 A discussion of the types of AP

APs can take many forms ranging from simple comparisons to sophisticated statistical models. As explained in the introduction it is convenient to categorise the procedures. The categories we adopt are those that use a single variable, those that use two variables and multivariate techniques. Single variable techniques are commonly those where a vector of values from a single variable are reviewed to identify something out of the ordinary. An example would be to compare this years travel and subsistence bill to that of last year. Such methods are implicit in the sense that there is no formal model relating the values. Two variable techniques allow the auditor to relate two pieces of different information. An example is the calculation of a ratio, which is then compared to other ratios. For example, it could be the ratio of bad-debt to revenue. In such a method the model formulation may be explicit or implicit. Frequently it will be implicit if standard ratios are compared automatically. For example, the audit manual of one of the large firms lists a number of ratios to aid the understanding of the balance sheet, e.g. fixed assets / intangible assets and current assets / liabilities. Multivariate methods require the auditor to build a model that explicitly acknowledges the relationship between the variables.

Other researchers (e.g. Fraser, Hatherley and Lin (1997)) have adopted another categorisation: of [1] scanning, [2] trend analysis, [3] ratio analysis and [4] modelling techniques. The last of these is split into reasonableness tests and statistically based methods such as regression. The two categories (by one, two or more variables or by type) are linked. Scanning and trend analysis use a single variable, ratio analysis normally uses two variables and modelling techniques, including bivariate techniques, are multivariate. The link between the two categorisations is discussed in more detail in section 3.5 and illustrated in Tables 3.13 and 3.14. The following paragraphs describe the categories. Clearly, the categorisations are used for convenience; there is an overlap between them. For example ratio analysis, using two variables might be used to compare in a convenient way parts of the profit and loss account (gross profit / cost of sales), but equally when using two or more variables this can be a modelling procedure.

3.4.1 Scanning

Scanning is a simple comparison procedure. For example, it might compare last year's account value to the unaudited value, using high-level data. For example it frequently compares annual data. The comparison could be measured as a percentage change, an absolute change in £, or a combination of both. It is a subjective evaluation based on knowledge of an audited organisation and past experience. A simple example of scanning is to compare one year's total pay bill with the previous years. The technique could be thought of as a natural consequence of the way many auditors set up their working papers. When doing the initial work, the 'current year' balances in the financial statements are put adjacent to the previous year's comparable figures. The purpose is to identify significant change, so that the change can be investigated. An anomaly might be that the difference between the value being tested and the expectation value is greater than some specified amount. Critical to this, is the judgement about what is a tolerable difference. Frequently differences of between 5% and 10% are set as the criteria. These ranges are arbitrary in that they are not explicitly linked to the overall materiality. They are based on past experience and common practice amongst auditors.

At its simplest the procedure involves a single variable compared between successive observable points, for example year 1 and year 2. A variant of the method is to scan a list of the balances (or transactions) to identify anomalies (outliers). Scanning is credited as being valuable as a pointer for further work. For a statistician, it is typical of something that should be done prior to the formal analysis of data. Chatfield (1985) discusses why this is a necessary prerequisite for analysis, and hence for APs. Implicit in the scanning technique is an assumption that there is a logical projection of the prior observation into the current year, that the two figures are comparable and have the same basis. Because of its superficial nature, the application of scanning may not take account of underlying changes in the organisation. It is often claimed by auditors (e.g. (AICPA 1998)) to be part of the next set of techniques: trend analysis. If the auditor is scanning many points this may be so, but two points do not provide a basis for a trend.

3.4.2 Trend Analysis

A collection of values occurring sequentially in time is known as a time series. One set of techniques used by auditors to examine such a series, at the planning, substantive and review audit phases, is trend analysis. The purpose is to analyse changes in account balances over a period of time. An example is the pay bill over time, the analysis being based on past observations of a single variable rather than some algorithm. Is the figure for pay, in the accounts currently being audited, compatible with the trend in pay over the last few years? The criteria for this assessment may be different for planning and substantive work. When considering the difference between their estimate, the expectation value, and the figure in the financial statements, the auditor, when doing a substantive test rather than using the AP as a planning aid, may apply more rigorous criteria for assessing whether the AP is signalling material error. The Audit Standard on APs, SAS 410, only refers to the accuracy of APs in the context of substantive APs. It says '*The extent of reliance that auditors place on the results of APs when used as substantive procedures may also depend on the following factors: a) other audit procedures directed towards the same financial statement assertions. For example, other procedures auditors undertake in reviewing the collectibility of debtors, such as the review of subsequent cash receipts, may confirm or dispel questions arising from the application of analytical procedures to an aged profile of customers' accounts; ... (b) the accuracy with which the expected results of APs can be predicted'*. The standards do not expect APs to be as robust at the planning stage: '*APs at this stage* (planning) *are usually based on interim financial information, budgets and management accounts'*. The standards acknowledge that the *'degree to which information can be disaggregated'* is a factor in the performance of an AP. This taken with the use of interim information implies that an AP such as simple trend analysis is more appropriate for the planning phase of an audit.

In trend analysis, the data being compared are financial, and trend analysis can take many forms. These are described using examples: simple trend analysis, graphical analysis, and average change over time analysis.

3.4.2.1 Simple Trend Analysis

Simple Trend analysis is a form of scanning that uses several years' data. For example, it might take the form set out in the Table 3.3.

Year	1990	1991	1992	1993	1994	1995	1996	1997
Pay Bill in millions of pounds:	12.5	12.9	13.5	13.9	14.6	15.1	16.0	16.9
Absolute Change (in millions)	0	0.4	0.6	0.4	0.7	0.5	0.9	0.9
As a percentage of 1990 value	100.0	103.2	108.0	111.2	116.8	120.8	128.0	135.6
Year on Year % Change		3.2	4.7	3.0	5.0	3.4	6.0	5.6

Table 3.3: An example showing three different APs based on trend analysis.

The first two rows document the data, the year and the pay bill. Suppose the year being audited is 1997 and the auditor has to assess the acceptability of the figure of £16.9m. Scanning the years 1996 and 1997 provides less information than reviewing the performance over the period since 1990. To get the best information from the data it is necessary to isolate changes such as a pay rise or staffing changes and apply that information to the audited figure of £16.9m for 1997. For trend analysis such considerations are informal; no explicit model is constructed. While the pay rises can be evaluated, the other variables such as staff changes are not taken into account in trend analysis. Methods in which the auditor takes account of more than one variable are considered as part of the modelling techniques such as reasonableness and statistical modelling in section 3.4.4.

Table 3.3 illustrates three different criteria for measuring the change: absolute change from year to year, percentage of a base figure (1990), and year on year percentage change. The criterion for a change that is material (sometimes known as 'tolerable difference', a concept that is described in Chapter 7) will be expressed differently for each of these. It might be an amount greater than £0.5m for absolute change or a year on year percentage change greater than 5%. In the example, the absolute change measure will flag a potential material error, since the £0.9m change is greater than the tolerable difference of £0.5m. Similarly, the year on year change also signals a potential problem.

3.4.2.2 Graphical Analysis

Another version of simple trend analysis is a review of the data through the inspection of a graph - *Graphical Analysis.* The same basic data used in the example at Table 3.3 form the basis of the example at Figure 3.4.

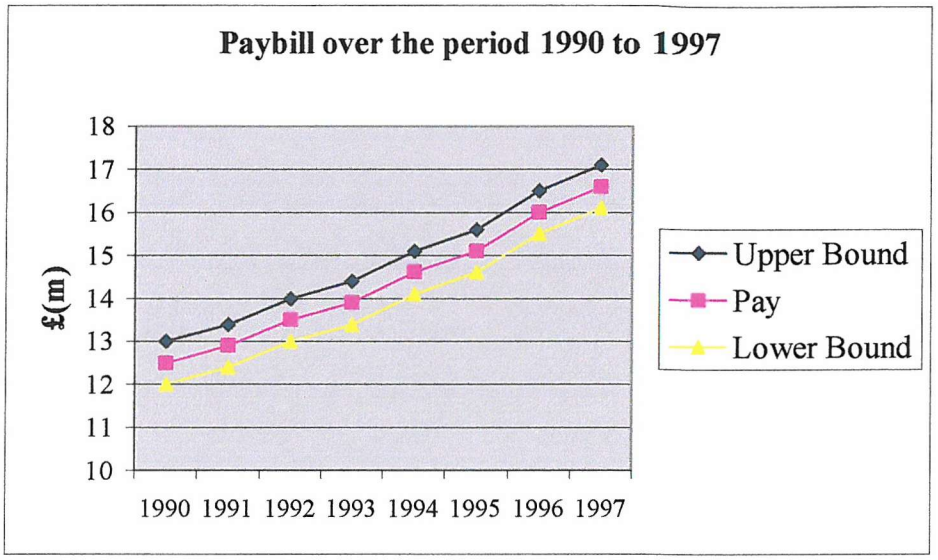


Figure 3.4: This graph shows the pay data plotted over the period 1990 to 1997. The 1997 figure is the projected value (£16.6m) to be compared to the value in the accounts. The graph indicates a steady increase in pay over time. The account value is £16.9m, which is more than the expectation value and below the upper bound of £17.1m. Note that the y-axis is scaled to start at 10. This was to make the graph more distinct.

Here the comparison is between some projected value and the value shown in the accounts together with a tolerable difference to recognise the closeness of the projected figure. Such a projection is done by eye in graphical analysis. A simple approach is to measure the difference and compare it to tolerable error. In the above example, based on the years 1990 to 1996, the predicted (projected) value for 1997 is £16.6m ± £0.5m. When compared to the value in the financial statements of £16.9m, this is within the 'acceptable' range. In some ways this tolerable difference is an informal version of the probability the statistician associates with the predicted value. If the observed value had been £18m, then that is clearly outside of the tolerable range of £16.1m to £17.1m, and would have signalled an unacceptable result and the potential requirement for a more detailed investigation.

If a large volume of data is available, it is preferable to use disaggregated data. This can take the form of breaking the data down into its components or more frequent

observations. For example, if there were four years' data, then forty-eight observations of monthly data are usually more informative than four observations of annual data. This is demonstrated by the data from a government agency for the period 1989 to 1993, as shown in Figures 3.5 and 3.6 below.

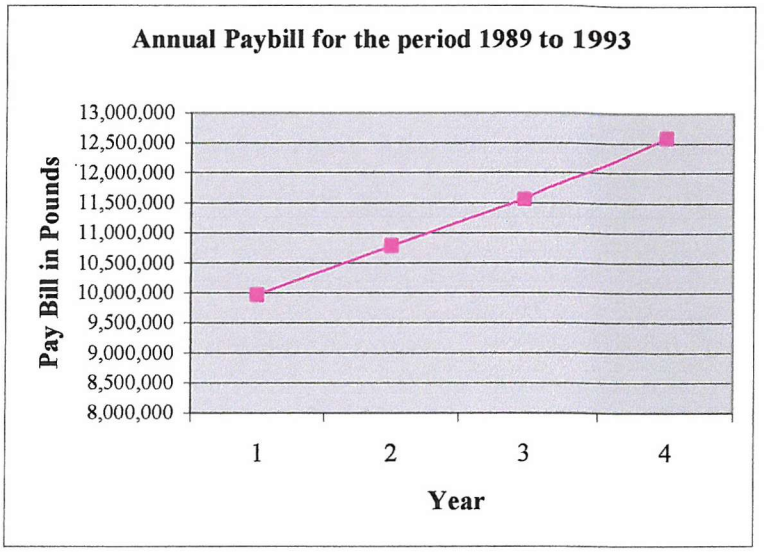


Figure 3.5. This graph shows the pay data for the agency over four years. Note that the y-axis commences at £8m.

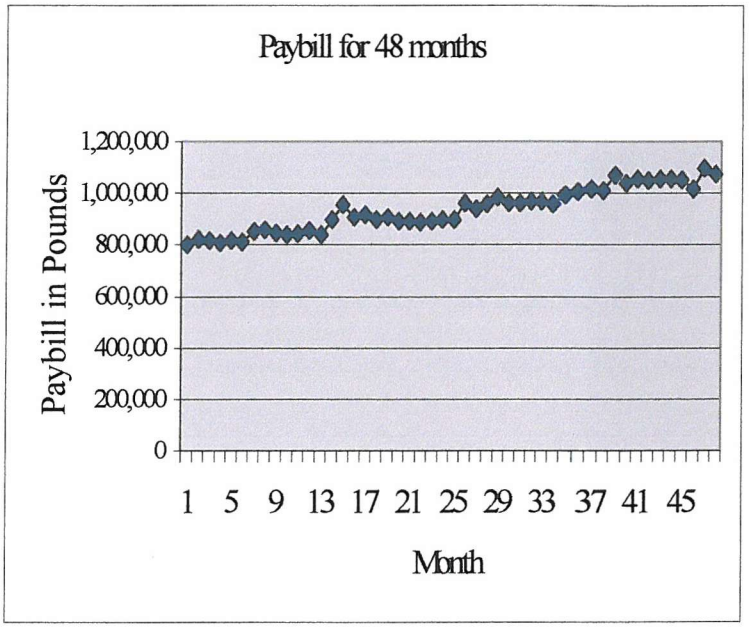


Figure 3.6. Pay data for the agency over four years (48 observations). It clearly shows the fluctuations in pay that the auditor may want to explore.

The graphs clearly show that looking at only the annual data would reveal a steady increase in the pay bill. The auditor might find this to be in agreement with his/her conceptions of the audited body. However, looking at the monthly data (figure 3.6)

reveals some dramatic changes from month to month around a steady growth in the pay bill. There is clearly more information here on which the auditor can base further investigations. Although the auditor is attesting that the annual financial statements are true and fair, they also wish to satisfy themselves that the underlying statements do not have compensating errors that might cause them to qualify their opinion. An example of this, although in a more general consideration, is the qualification to the Accounts of the Department of Social Security. There for many years the NAO put an estimate on the total value of error (overpayments, frauds, etc.). The net value was not large, but the compensating over and underpayments were indicative of something being wrong with the financial statements.

3.4.2.3 Average change over time analysis

Using the data in figure 3.3, the calculation of a*verage change over time analysis* can be used to predict the pay for the year being audited. First calculate the average change for the period 1990 to 1996 (7 observations). This would be:

$$\frac{\text{Pay in 1996 - pay in 1990}}{\text{(Number of observations - 1)}}$$

$$= \frac{16.0 - 12.5}{(7-1)}$$

$$= £0.58\text{m}$$

So the prediction for 1997 is the value for 1996 + average change per year = £16.0m + £0.58m = £16.58m. When comparing this to the financial statement figure of £16.9m and allowing for the tolerable error of £0.5m, the predicted value is said to be 'acceptable'.

Trend analysis is appropriate when a relationship is stable, but less so when there are operational or regulatory changes. The process is heuristic, the development of expectations being implicit. The number of years required for trend analysis to be appropriate is a function of stability of the environment of the AP. Because of the unstructured nature of the process, it is essential that the auditor understands how the environment of the audited body changes, both internally and externally. The example illustrated by Figures 3.5 and 3.6 demonstrates that trend analysis is likely to be more effective if disaggregated data is used.

Chatfield (1975) provided what he called a 'loose description' of trend analysis. It is one that fits perfectly here: '*long term change in the mean*'. This raises the question of what is long term? That must be a subjective judgement informed by some knowledge of the data: for example, taking account of cycles in the data. As will be seen from the surveys described in Chapter 5, the auditors' application of trend analysis implies that the prediction is exact: a deterministic model, where the future is exactly determined by the past. Although their models are representing accounting identities, they are incomplete because their interpretation and knowledge of the audited organisation is incomplete. Thus for most implementations of trend analysis, the model is not exactly determined by the past. Such series are stochastic and an exact prediction is impossible. A range of other potential values that could arise from the same data series surrounds the predicted value. These potential values have a probability distribution that is conditioned by past values. The task is to infer that range of values.

3.4.3 Analysis using Ratios

Ratio analysis looks at the relationship between values: put simply it transforms two variables into a single number. The ratio might involve the relationship between the sales of a shop and the floor area devoted to sales. The ratio will report a single number: the sales per square metre. In this simple definition there is no restriction on the units of either variable. Financial ratios are a subset of ratios and consider the relationship between values in the financial statements or between values in those statements and the accounting systems underlying those statements. These are the ratios we consider in this section.

Ratios can be traced back to the Greeks: Euclid analysed the properties of ratios as part of his work on geometry. The first modern use of ratios can be traced back to the time of the Civil War in the USA, when the increase in US bank credit gave rise to the segregation of current and non-current items and the development of the ratio of current assets to current liabilities. Another landmark was the du Pont Company's ratio 'triangle system' to evaluate its operating results (1919). These are examples of the normative use of ratios, which summarises the relationship between Y and X in a single number, which can then be compared to some standard. That standard can be theoretical, or based on past experience of the organisation being audited or past experience of other organisations (e.g. Trade Associations) or by specific enquiry.

One application of normative ratios could occur at the planning phase of the audit, where they might be used to highlight areas that are performing differently from the norm. The areas might be branches of a chain of shops, social-security offices or railway stations on the underground. Different in this sense is not necessarily bad, but worth investigating.

In addition to the normative role for ratios, they are also used in a predictive way. In their predictive role, the ratios are part of the estimation of a functional relationship, usually for the purposes of prediction. Here they might, for example, seek to predict the future profit of a company by using the ratio of profit to sales. In his paper reviewing the use of, and analysis of, Financial Ratios, Barnes (1987) expanded Whittington's (1980) description of these predictive uses. Barnes (1987) defined two main users of predictive ratios: '*The positive use of ratios has been of two types: by accountants and analysts to forecast future financial variables, e.g. estimated future profit by multiplying predicted sales by the profit margin (the profit/sales ratio), and more recently by researchers in statistical models for mainly predictive purposes such as corporate failure, credit ratings the assessment of risk, and the testing of hypotheses in which inputs are financial ratios*'. The first of these would include accountants in their role as auditors.

The use of ratios as a management tool has a long history. Accountants and economists have advocated them over the years. Auditors are trained accountants, so the transition from the management role to an audit one was a natural one. To the internal manager, whatever the theoretical grounding of ratio analysis, there is a practical consideration: do the ratios provide the user with valuable information? Deakin (1976) captured the essence of this point '*Thus, the ultimate test of the value of such a model (ratio) lies not in its adherence to certain data assumptions, but, rather, in its adherence to its usefulness in decision-making*'. That argument is attractive to the manager of a business whose decisions affect his own business: an internal matter.

Unlike the manager, whose role was to serve the internal management, the auditors have a different audience: the public in the form of the shareholders or the wider public such as those who might act on the financial statements. In the public sector it is Parliament and the public at large. It is the auditor's role to attest to the truth and fairness of the financial statements. Their use of ratios (or any procedure) can affect the

decisions of others; so they have an implied responsibility to ensure that their work is reliable and can be shown to be so. To be defensible, the use of ratios as an AP should be consistent with the assumptions underlying the method, both implicit and explicit. Whittington (1980) argued that those using ratio analysis should understand and be aware of the assumptions that underlie their use. He put it this way: '*There is also a growing body of empirical literature which studies the statistical distributions of, inter-relations between, and the predictive content of accounting ratios, but in this literature also, it is customary to assume that the ratio is the appropriate statistical form for summarising the data, without explaining what assumptions are necessary for this to be the case*' Whittington (1980). This debate about the use of ratios has been going on for many years. In 1968, Altman alluded to the differences in the consideration of appropriateness of the use of ratios between academicians and practitioners. He wrote '*Can we bridge the gap, rather than sever the link, between traditional ratio 'analysis' and the more rigorous statistical techniques which have become popular among academicians in recent years?*' (Altman 1968). He was alluding to the application of ratios without consideration of the appropriateness of a ratio in a particular set of circumstances. To apply a standard range of ratios without consideration of the circumstances and whether they are theoretically correct can lead to erroneous conclusions.

Such an application would have analogies to the application of IDA techniques by a statistician. Chatfield (1985) described such a strategy in his paper on Initial Data Analysis (IDA). There the statistician will explore the data to understand it better without worrying about the theoretical appropriateness. But they do not draw conclusions from that: it is purely exploratory.

The Auditing Standards encourage the use of ratios. In Practice Note 13 (The Audit of Small Businesses) they give an example of a type of AP that uses a ratio. The Practice Note says: *'The use of widely recognised trade ratios (such as profit margins for different types of retail businesses) can often be effectively used in APs to provide evidence to support the reasonableness of recorded items.'*

A survey of auditors reported by Daroca and Holder (1985) found that over 50% of auditors used ratio analysis. Three of the most common ratios were current ratio (e.g. current assets/current liabilities), gross margin on sales and profit margin on sales. In

ratio analysis the comparison may be over time, either within a particular company or between organisations. The former is *time series analysis*, also known as *longitudinal* analysis. For example, the ratio might be sales as a percentage of capital, and the analysis comparison with last year's ratio. Indeed ratio analysis could be part of trend analysis if ratios for several years were used. In contrast, *cross-sectional analysis* compares the same ratio, at the same time period, across many similar companies and/or to some industry norm. For example, the comparison may be between the sales per square foot of Safeway's and Sainsbury's, or may be measured against some industry/business norm (comparing the sales per square foot to the 'average' of all supermarkets).

If both the numerator and denominator are in the same units, as they are in financial ratios, the ratio has the potential to transform the variables to a relative value and so compensate for the effects of size or changes in purchasing power: *the value of money*. Figure 3.7 below shows the ratio of stock to sales for 16 companies in the same industry category (brewing) for one particular year: 1991. The comparisons were of companies of very different sizes from small family firms to large multi-national companies. The sales ranged from £1.75m to £1,461m. It can be seen that most companies have a comparable ratio: within the range 0.04 to 0.10.

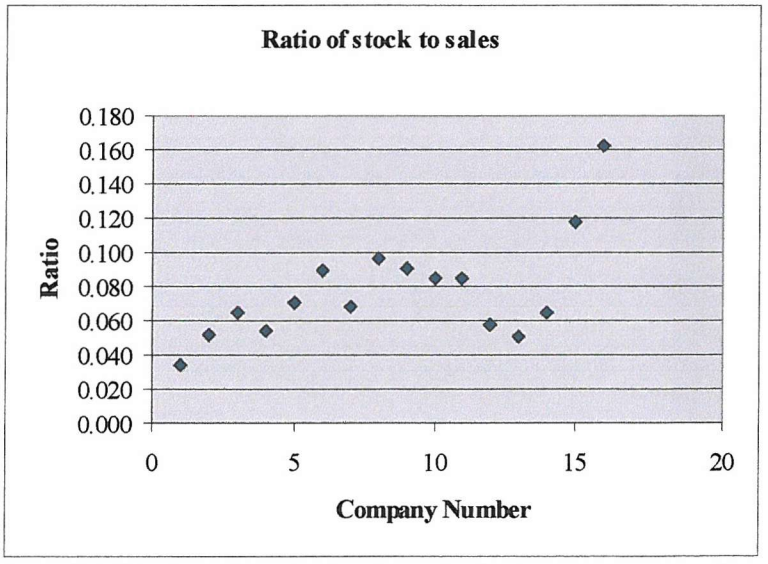


Figure 3.7: This shows the ratio of stock to sales for the year 1991. It can be seen that the ratio varies over a small range for most companies but there are also wide variations. [Source: Compiled from an analysis company accounts of Brewing Companies undertaken for this thesis].

However, the interpretation of such ratios can be complex because, for example, of the nature of the businesses. They are all from the brewing industry, that classification being based on the FT Extel database. The information in Figure 3.7 is from that database. But in fact investigations revealed that some are also distillers and/or are in the catering trade. One company that was rejected from the analysis had quit the brewing business and now ran homes for the elderly.

The full range of values is made clear when inspecting the data in Table 3.8. It transpires that companies 18 and 19, although they are classified as brewing, are also major distillers. The stock holding for such companies is entirely different; brandy is aged for years, whereas many beers have a short shelf life. The stock patterns are very different. The way things can dramatically change between years is illustrated by results for company 1: inspection of Table 3.8 makes it is clear that some major change occurred between 1993 and 1994. In fact further investigation revealed that there had been a takeover that increased the stock and sales.

	Stock (£m)				Sales (£m)			
	1991	**1992**	**1993**	**1994**	**1991**	**1992**	**1993**	**1994**
1	0.06	0.03	0.02	3.04	1.75	1.22	0.64	22.93
2	0.60	0.63	0.64	0.52	11.65	12.54	12.65	12.11
3	0.71	0.66	0.88	0.82	10.97	13.31	13.79	17.74
5	1.21	1.49	1.26	1.09	22.54	25.74	28.03	28.09
6	1.20	1.18	1.72	1.54	16.94	18.61	28.07	33.07
7	2.41	2.46	3.16	3.04	27.04	29.58	32.20	32.07
8	1.94	1.71	0.99	1.15	28.21	32.12	34.52	35.94
9	3.02	3.57	2.40	2.61	31.34	35.82	39.11	43.81
10	3.80	3.39	3.27	3.60	42.17	43.07	42.73	42.15
11	4.99	5.31	5.02	4.06	58.52	68.14	72.90	72.34
13	10.70	11.98	11.19	9.07	126.26	128.58	62.55	61.00
14	5.12	6.53	2.81	2.79	88.32	98.33	102.94	104.21
16	24.54	35.61	46.64	40.32	487.68	529.06	596.05	720.36
17	54.66	86.01	50.86	49.05	844.83	1,333.15	1,326.13	1310.20
18	936.00	902.00	1,031.00	859.00	7,952.00	8,798.00	8,892.00	8,709.00
19	2,370.00	2,490.00	1,830.00	1,820.00	14,610.00	15,730.00	16,140.00	17,110.00

Table 3.8: The Value of Sales and Stock for 16 different companies for the year 1991-1994. The units are **£m**. The blank lines are firms excluded here because there were missing data in one or more years. It will be seen that there is a wide variation in size. [Source: Analysis of company accounts of Brewing Companies undertaken for this thesis].

Thus interpretation is an issue, particularly in cross-sectional analysis. As demonstrated by this example, seemingly comparable industries may have very different

characteristics. Salmi, Dahlstedt, Luoma and Laakkonen (1988) looked at the possibility of using ratios to predict the industry classification of a firm. This is the reverse of the problem discussed in this thesis, but an interesting one. They concluded, *'it is probably the industry classification which ought to be reconsidered rather than the soundness of financial ratios'*. This makes the point that the choice of comparators is one that should be done carefully. The accepted industry classification may not be appropriate.

The example given above illustrates one of the basic assumptions underlying ratios. It is that of proportionality with the basic form $x/y = \beta$, or does it involve some constant, α. In the above example, it is because there is a basic underlying stock holding, the relationship between x and y might be $y = \alpha + \beta x + \varepsilon$. To obtain the ratio we divide by x to give: $y/x = \alpha/x + \beta + \varepsilon/x$. For proportionality we want $\alpha = 0$. If $\alpha \neq 0$, then y/x the proportionality assumption does not hold.

In the example, industries had very different components for their business. We can illustrate the problem with a simplified example, ignoring the associated beverages and snacks that the brewers dispense. Suppose it was initially thought that only beers made up the sales, and suppose the sales and stock were y and x respectively. Suppose the model was thought to be $y = \alpha + \beta x + \varepsilon$, but then it was discovered that there were sales (y), stocks of beer (x_1), spirits (x_2) and food (x_3). So the model is now more complex: $y = \alpha + \beta x_1 + \gamma x_2 + \delta x_3 + \varepsilon$. The question is: what is the ratio. Should one use a complicated model, or three simple models such as: $y_1 = \alpha_1 + \beta x_1 + \varepsilon_1$, $y_2 = \alpha_2 + \gamma x_2 + \varepsilon_2$, and $y_3 = \alpha_3 + \delta x_3 + \varepsilon_3$. This simple example illustrates that it might be necessary to disaggregate the analysis and hence the data to an appropriate level.

A distinction between trend analysis (described in 3.4.2) and ratio analysis is that trend-analysis only looks at the behaviour through time of the data for one account figure, while ratio analysis measures the relationship between different data sets. Given the inter-relations between the account figures, one might expect ratio analysis to be more likely to give consistent measures or ones that are not subject to sudden changes unless there is a change in the underlying relationship. For example, it might be that as sales of goods increase, so the cost of materials used to produce those goods would increase in proportion. If the relationship is linear, then simple ratio analysis has the potential to be a useful analytical tool. While a single account figure can vary for a number of reasons,

an assumption commonly made in using ratio analysis is that there is a stable relationship between the chosen account figures. A variation in a ratio may be a signal of a change in circumstances of the business or of a potential problem (e.g. error or fraud). The development of the expectation is implicit, and is a function of stability of the environment in which the AP is applied, and the data available for the analysis. This area will be picked up and discussed later in Chapter 7: *Investigations of whether APs correctly signal the presence or absence of error*. In particular we review some work by Berry and Nix (1991), who did some interesting work comparing simple ratio cross-sectional analysis to regression based analysis.

3.4.4 Approaches that explicitly use many variables to develop an Expectation

Trend analysis described in 3.4.2 looked at the performance of a variable over time. One way of conceptualising the time variable is as a proxy for other variables. Time might be used instead of changes in the market, caused by the growth of the organisation. It is simpler and easier to conceptualise than the inter-related changes of interest rates, market changes and other economic variables. When considering the expectation value for pay over time in trend analysis, the auditor could think of other things that influence the pay figure. But if the auditor explicitly used the variables such as staff numbers, staff changes and pay, then the technique can be categorised as modelling. Similar arguments to those about disaggregated data indicate that the use of the variables rather than the time proxy should give more precise and reliable expectation values.

APs that use multivariate data and are characterised as 'modelling' are distinguished by the identification of meaningful, stable relationships between financial and non-financial data. Two examples of this approach are *reasonableness tests* and *statistical modelling* (such as ordinary least squares regression, or nonparametric regression.) The data may be from within the audited body, external to it or a mixture of the two.

3.4.4.1 Reasonableness Analysis

Reasonableness analysis is a structured model that attempts to estimate the expectation value using information that the auditor considers relevant to that expectation. Such procedures are considered by auditors to allow them to use their intuitive judgement to play a major role. An example is where the auditor uses data on the number and

categories of staff, the number of hours, days or weeks in the period and the relevant average pay rates to predict the figure for salaries and wages in the accounts. Another example is the calculation of an expectation value for income of the Passport Agency based on number of passports issued and the standard charges by passport type. If the data were known exactly, and there were fixed rates, then this would be a deterministic model. For example if there were three types of passport, all strictly controlled because of the inherent security issues around passports, and there were a fixed rate for each passport type, then it should be possible to calculate exactly the income from fees. This has been successfully carried out in the audit of the UK Passport Agency. Other examples include the calculation of an expectation value for hotel income based on occupancy rate of hotel rooms and the average room rate (possibly by room type), and the calculation of the Driver and Vehicle Licensing Agency (DVLA) income using the number of licences issued and their fee rates.

How a reasonableness test might be used can be illustrated by an example, based on the data from the example in Annex A on page 67. If the auditor knew the number of staff and the average pay from reliable sources then a simple expectation value could be built. The total pay might be the sum of the average number of staff in each group times the average pay for that group, after allowing for the effect of inflation. The details are set out in Figure 3.9 below.

Details of the Reasonableness Test	
Average staff for audited year	849 .63
Average pay from last year	13461 .89
Allow for inflation of 3.18%	13889 .98
Expectation value for pay in the year1998-1999	11801272 .64
Pay figure in the accounts	12597333 .00
Difference	-796060 .36
Difference as a percentage of expectation value	**-6.75%**

Figure 3.9: An example showing the results of a reasonableness test on payroll.

This might be an unacceptable difference. Suppose now that the auditor realises that pay inflation was not 3.18% but 4.5%, would that make a difference? The effect of this is shown in Figure 3.10; again, the difference between the expectation value and the book value is not small. This crucial subject of acceptable and unacceptable differences will be examined in Chapter 8.

Details of the Reasonableness Test	
Average staff for audited year	849 .63
Average pay from last year	13461 .89
Allow for inflation of 4.5%	14069 .57
Expectation value for pay in the year1998-1999	11953857 .34
Pay figure in the accounts	12597333 .00
Difference	-643475 .66
Difference as a percentage of expectation value	**-5.38%**

Figure 3.10: This shows the results of the revised reasonableness test on payroll.

The effectiveness of these APs will depend on the diligence of the auditor and the quality of the data. Whether the procedure is cost-effective depends on its ability to reduce the costs of doing the audit, while still meeting the standards in terms of accuracy and achieved assurance (efficiency criterion) and whether their application is capable of producing a result (effectiveness criterion). Reasonableness analysis could be described as an informal version of statistical analysis, using financial and/or non-financial data to predict an account figure.

3.4.4.2 Statistical Modelling

In statistical analysis, as in reasonableness testing, the auditor develops a model to predict, or explain, the figure in the account (the response variable) using relevant data, both financial and non-financial data (explanatory variable(s)). However, in contrast to the deterministic relationship underlying the reasonableness test technique, statistical models are based on derived statistical relationships between the explanatory variables and the response variable. Both techniques require the user to define variables, between which there are expected to be meaningful relationships. An essential difference is that statistical modelling explicitly recognises that the expectation value is not exact, but that it has a probability distribution that is conditioned by the values of explanatory variable(s). It is possible to estimate the precision of the result and the probability that the response variable lies in any given range. Thus the precision of the auditors' expectation can be quantified. Statistical modelling is not just a one-to-one pairing of response and explanatory variables. It tries to assess the relative importance of different explanatory variables, and can thus extract the significance of competing or complementary variables.

Statistical modelling could be used to predict the pay expenditure of an organisation. It might use as explanatory variables: staff numbers, promotions, and details of starters and leavers. Another example is an explicit expectation for income of an ice-cream manufacturer based on temperature, day of the week and other variables known to affect sales. Examples of other potential relationships between response variable (financial) and explanatory variables (usually a mixture of financial and non-financial) are set out in Table 3.11 below.

Response Variable	Explanatory Variable
Unemployment Benefit	Labour statistics
Subsidy	Number of cows
Child Benefit Expenditure	Number of children in certain age groups derived from the census and updated since the last census

Table 3.11: Examples of the relationships between variables.

An example of an AP based on statistical modelling is illustrated in the example set out in Annex A on page 67. Essentially, there are details of two staff groups with the data being on a monthly basis. Associated with the data are details of the total pay, also on a monthly basis. A simple regression model can be built using the staffing and pay data for the 36 months prior to the current audited period. The model is:

$Y = 913680.5 + 7239.8 \times V_1 - 3394.9 \times V_2$, where Y is the expectation value, V_1 is the variable 'higher grade staff' and V_2 is the variable 'junior grade staff. That model is then applied to the staff data for the audited period to produce twelve expectation values, one for each month. These are combined to produce the expectation value for the year. The results are in Figure 3.12. For more detail see Annex A on page 67.

Account Value	Expectation Value	Difference	Difference as %
£12,597,333	£12,341,780	£255,553	2.09%

Figure 3.12: The results of regression based AP on payroll. [Source: Based on data from a Government Agency].

3.5 The relationships between the Techniques

The Auditing Standards in SAS 410 describe the nature (their term) of APs to include comparisons and relationships. For comparisons, their examples include the consideration of the entity's financial information in relation to comparable information from prior periods. Another example that is given is the comparison of a particular ratio with the relevant industry average. The description in the Standards of '*comparisons*' is

compatible with the one / two variable(s) type APs such as scanning, trend analysis and simple ratio analysis.

The Auditing Standards then go on to describe another type of AP: '*the consideration of relationships between elements of financial information that are expected to conform to a predicted pattern....*'. Examples they provide include relationships of gross profit to sales, or between *'financial and relevant non-financial information'*. Clearly, these APs are based on two or more variables. Some form of modelling is involved, be it the intuitive reasonableness tests or the more formal statistical type procedures.

The standards are describing a range of methods that make up the universe of all APs. The different methods are often described as distinct; they are not! Consider ratio analysis, which Blocher and Patterson (1996) defined as the comparison of two variables, both of which are usually financial. Others have a wider definition where ratios are based on financial and non-financial data, for example, sales per retail square foot. A ratio is sometimes considered to involve a function of several variables. For example, sales per square foot over time. Here the AP has moved from a comparison to a consideration of the relationship between different sets of data and becomes a modelling procedure. When comparing two variables it may be described as ratio analysis, but if three or more variables are involved then it is a modelling procedure. That might be either informal modelling such as reasonableness tests or the use of regression (or some other statistical procedures). This is not just semantics; it reflects the type of expectation value moving from implicit to explicit.

The Auditing Standards recognise the distinction between two types of APs. This and our interpretation are summarised in Table 3.13. The five types of AP can be mapped onto the two categories defined in the UK Auditing Standards. This mapping is shown in Table 3.13. The research work described in the literature review with regard to the five categories can be translated into comparison or relationships as set out in the UK guidelines. The concepts are consistent with the categorisation by the number of variables, which we prefer.

Categorisation in the Auditing Standards	Categorisation by Method	Categorisation by Number of Variables
Comparison	Scanning	Single
	Simple Trend	Single
	Ratio	Two
Relationships	Ratio	Two or more
	Reasonableness	Multivariate
	Regression	Multivariate

Table 3.13: The connection between the description of APs in the Auditing Standards and in the literature.

The data used for the AP can be from a variety of sources: financial and non-financial data; the latter, if information is available, has the potential for greater power in producing defensible audit statements. That is because it is from different systems to those producing the financial statements. Both financial and non-financial data can be subdivided into data internal or external to the organisation. Because data from outside of the organisation are independent of the accounting and other systems of the audited body, they have the power to be particularly useful in developing the expectations for income. Of course, such data can also be indefensible if grounded in a different context. However, they can sometimes be used to crosscheck the results of an AP. The Auditing Standards stipulate that: '*The extent of reliance that auditors place on the results of APs when used as substantive procedures may also depend on the following factors: (a) other audit procedures directed towards the same financial statement assertions*' (SAS 410). A potential example involves the audit of Child Benefit paid by the Child Benefit Agency. Suppose data on child numbers could be obtained from a different Government agency to the Child Benefit Agency, for example the Department of Education. In addition, if information on family size was available from outside of the Child Benefit Agency, then a benefit that was based on children could be predicted independently of the Child Benefit Agency systems.

Scanning and trend analysis, because they use only a single variable to predict, do not use external or non-financial data. Modelling procedures are able to use and combine all four sources of data, financial and non-financial, internal and external, thereby increasing the power of the techniques. Table 3.14 below summarises these relationships.

Category of AP	Explicit / Implicit	Number of variables used to build the expectation	Non-financial data	External data	Quantified precision.
Scanning	Implicit	One	No	No	No
Trend Analysis	Implicit	One	No	No	No
Ratio Analysis	Implicit	Two	Yes	Possible	No
Reasonableness	Explicit	Two or more	Yes	Yes	No
Regression	Explicit	Two or more	Yes	Yes	Yes

Table 3.14: This summarises the properties of the different classes of AP. (Source: Adapted from Audit Practice Release – APs (AICPA 1998)).

As the role of APs changes and the expectations developed become more explicit, the power of the AP also changes. Libby (1981) addresses this issue, which is considered in the Literature review in Chapter 4. A way that bridges the distinction between the categories is to define the methods as moving from one category to another as the objectives of the procedure change or the data used change. This can be illustrated by another example. Simple trend analysis as described above uses one variable over time. It is an informal procedure in the sense that it may be graphical with the trend modelled by drawing a line through the data points by eye, or use one of the other methods described in section 3.4.2. It is also informal because the expectation value is not derived explicitly; it is implicit. If two or more variables are used then the method may become a reasonableness test where the formation of an expectation value is explicit; indeed regression could be used, if it were appropriate. Table 3.14 summarises this information.

3.6 Discussion of implications for the application of an AP

The issues discussed above are not academic in the sense of being irrelevant to practice. The American Institute of Certified Public Accountants (1998), in guidance on the use of APs, said, *'The effectiveness is a function of precision of the expectation developed by the auditor. The precision of the expectation can be increased by forming expectations that take into account an understanding of the client's industry and the factors affecting the industry'*. The AICPA then cite several frauds that they consider should have been initially identified using APs, illustrating the importance of proper consideration of the expectation. The frauds are listed in Annex B on page 74.

We now consider in more detail the outcomes of an AP and the rules that underlie the development and subsequent analysis of the expectation-value.

3.7 Outcomes of an AP

As described in Chapter One outcomes from the application of a substantive AP can be divided into four categories according as (i) material error is or is not in fact present, and (ii) the AP signals that material error is or is not present. The four categories are shown in Figure 3.15.

		True Position	
		No Material error present	**Material error present**
Audit Result	**AP signals no material error**	Correct signal	False-positive signal
	AP signals material error	False-negative signal	Correct signal

Figure 3.15: Possible outcomes from the application of an AP. [Source: This thesis].

3.8 A structure for APs

APs are based on a variety of methods, ranging from statistical to qualitative. If we start from the premise that they are all useful and valuable, then some framework is needed that demonstrates to third parties that the analysis is valid in two important respects. Firstly, that it can be relied upon to produce evidence that there is no material error, where this is in fact the case. Secondly, if there is material error the AP can be relied upon to indicate the presence of such error, although other audit work may be required to resolve and provide the conclusive proof. The latter work is not addressed in this thesis; for our analysis the criterion is that the AP must be a reliable and consistent indicator of the presence or absence of error.

In order to bring a structure to the consideration of APs, we introduce the concept that an AP can be thought of as two rules: an investigation rule and an outcome-rule. The investigation rule is about building the AP model and calculating the expectation value. The outcome-rule is concerned with assessing the closeness of the expectation value to the book value that is being audited. The flow diagram set out in Annex C on page 76 is a schematic representation of the AP process. The basic structure is forming an expectation about outcome, the expectation-value, and then making a comparison between the audited value and the expectation-value and making a judgement about the

difference. The components that make up the investigation rule are coloured green in Annex C, and those of the outcome rule are coloured turquoise.

As we pointed out in section 3.2.3, unlike the Auditing Standards in the USA that provide some structure for the process, the UK Standards do not. Thus the structure proposed here addresses that omission.

3.8.1 Investigation Rule

The investigation-rule is the initial step in developing an expectation-value for the figure in the financial statements. It seeks to incorporate causal links between different sets of data and utilises the auditor's understanding of the plausible relationships when using procedures such as ratio, trend or modelling analysis. The auditor's understanding and knowledge of the client's business are an essential part of this process. The Statement of Auditing Standards on APs states that: '*The application of APs is based on the expectation that relationships between data exist and continue in the absence of known conditions to the contrary*'(SAS 410), clearly recognising that an understanding of the inter-relationships is a key part of APs. Section 3.4 looked at a number of different categories of AP. The auditor chooses the one that meets the objective of their audit test. The investigation-rule produces the expectation-value: $\hat{Y}_T$. The other values of interest are the book value [Y_B] and the true value [Y_T], which is not known, but which it is hoped will be close to the value in the accounts. The relationship between $\hat{Y}_T$ and Y_T is $E(\hat{Y}_T) = Y_T$.

There are the two purposes for using an AP: to explore and explain the relationships between the two or more variables, for example response and explanatory variable(s), and secondly to predict a future value. The exploration of relationships is especially important at the planning stage of an audit. Since this thesis is particularly concerned with the predictive power of APs, it concentrates on aspects of prediction. Our approach will be to build an investigation rule to summarise the data, and then use this rule to explore the relationship(s) and/or to predict a future value.

The predictions (investigation rules) may involve cross-sectional or time-series analysis. The former may involve comparison between different parts of an organisation, or between similar parts of a number of organisations: the parts could be locations or branches within an organisation. Alternatively, the cross-sectional analysis

could involve particular feature, for example creditors, from many different organisations. It does not involve the passing of time. For example if the Passport Agency had many different locations, the predictions could be used to measure those that were performing differently in some way to the others: a particular location may be better, or worse, by a specified amount. Another example of cross-sectional analysis could be the use of data about the ratio of sales to stocks for a number of similar organisations.

An alternative analysis might be to look at the data over time, say the number of new passports processed each month and the expected income from them. This is analysis over time, otherwise known as longitudinal analysis. The rest of the thesis will explore APs that deal with data over time.

3.8.2 The outcome rule

The outcome-rule is the other key part of an AP. It is concerned with assessing the outcome of the investigation-rule. The outcome-rule may be arbitrary or statistically based. It would be arbitrary if the rule is set by judgement and might, or might not, be linked to the overall materiality.

Once the investigation rule has produced the expectation-value, $\hat{Y}_T$, we need some criteria to measure how close it is to the value being audited Y_B (the book value). That is, we need a rule for assessing the closeness of $\hat{Y}_T$ and Y_B. However there are the three values we defined above: Y_B, $\hat{Y}_T$ and Y_T and the relationship between them can be described in the following way. It makes clear the components of the prediction. The observed difference between the Expectation Value and Account Value is $\hat{Y}_T - Y_B$. Adding and subtracting the true value gives: $(\hat{Y}_T - Y_B) = (\hat{Y}_T - Y_T) + (Y_T - Y_B)$. This can be interpreted as: Observed Difference = Prediction Error + Target Difference. The auditor wishes to measure the Target Difference, but observes $(\hat{Y}_T - Y_B)$. Only if one knows the true value of the population will Y_T be known. The auditor should be aware of the prediction error. Blocher and Patterson (1996), who drafted much of the current guidance on APs in the USA recommend that one of the key steps in evaluating an AP is an assessment of prediction error.

A statistically based AP will try to estimate the prediction error. For example: if linear regression were used to estimate Y_T, then the associated confidence interval takes account of the potential prediction error and is:

$$\hat{Y}_T \pm t_{\alpha/2,\, n-2}\; s\; \{1 + 1/n + (x_i - xbar)^2/\Sigma(x_j - xbar)^2\}^{\frac{1}{2}}$$

where

n	=	number of observations in the prior period
j	=	1....n
$(1 - \alpha)$	is	probability associated with the prediction interval
s	is	standard deviation
xbar	is	mean value of prior data
x_i	is	value of explanatory variable at value being predicted
x_j	is	j^{th} value of the prior data

We will return to the prediction error later, but for now we will concentrate on the observed difference.

We define $|\hat{Y}_T - Y_B| \leq C$ as the outcome-rule, where C is some specified value, which we will call the outcome-value. Auditors have a value they use as the specific value 'C'. It could be materiality, or a simple function of it. We discussed materiality in Chapter 2 where we saw that the Auditing Standards require that materiality should be taken into consideration in assessing the audit of particular values within the financial statements. A relevant extract from the Standards is: '*Materiality may also be considered in the context of any individual primary statement within the financial statements or of individual items included in them*' (Glossary of terms in the Statements of Auditing Standards). Because the term materiality is used in different ways to describe the assessment of the outcome of the overall audit, or of part of an audit, we will call the specific value for APs the 'outcome-value' to distinguish it from the auditor's 'materiality'.

This link between the materiality figures at the higher level (overall account) and the outcome-value applied to an individual line item within the financial statements is important because auditors claim to work to 'tight' materialities (e.g.1% or much less). The question is whether that means they work to tight materialities relative to the overall level of the financial statements or to the individual tests that together make up the overall audit opinion. The possible implications of this can be illustrated by an example.

Suppose for the audit of the UK Passport Agency's financial statements (the example in Chapter 2) the auditor defines materiality using the total income in the financial statements as a yardstick, to be £1.075m (1% of an important value (income) in the financial statements, rounded to 3 decimal places[1]). Suppose also that an AP is being applied to the 'amendments' income figure: whose value within the financial statements is £17m. It may be one of many lines, or balances, within the financial statements, each of which may be subjected to a separate substantive test. For this example there are two such lines, or balances, that make up the total of the financial statements. Table 3.16 shows the alternative effects of applying the overall materiality to be the outcome-value for each line item in the set of financial statements and the outcome-values if they were set at 1% for each line of the financial statements. Outcome-value 1 shows the outcome-value as a percentage of the income figure if it is set at overall materiality, while outcome-value2 shows the value it would be if it were set at 1% of the line, or balance, being tested.

Income Values in the Financial Statements	Values	Overall Materiality	Outcome-value 1 (overall materiality)	Outcome-value 2 (% of balance)
Issue of passports	£90,540,920.00	£1,075,409.20	1.2%	£905,409.20
Amendments (change of name or photograph)	£17,000,000.00	£1,075,409.20	6.3%	£170,000.00
Total	£107,540,920.00	£1,075,409.20	1.0%	£1,075,409.20

Table 3.16: A comparison of setting the outcome-value for line items based the overall materiality and contrasting them with the outcome-values if they were set at 1% for each line of the financial statements. [Source: Based on the accounts of Passports Agency].

Which outcome-value should be used for the outcome-rule for the application of an AP to the 'Amendments' line in the financial statements? If the outcome-value used is the overall materiality figure of £1.075m, it represents 6.3% of the balance being audited for the 'Amendments' line item. Alternatively applying an outcome-value that is 1% of the line item gives a value of £0.170m. We thus have two very different criteria for measuring the outcome of the AP and whether the expectation value is close to the book value of £17m for the 'Amendments' line in Table 3.16.

An issue for this thesis is how materiality translates into 'outcome-value', especially when materialities as low as 1% to 2% are applied. Loebbecke and Steinbart (1987) found that a much higher percentage, typically of the order of 10% is frequently applied

[1] In real life the auditor would normally round the figure to convenient whole number such as £1,000,000 in this case.

for the outcome-value. We will return to this in Chapter 8, when considering the performance of the APs against different levels of outcome-value / materiality. That performance will be assessed in terms of the level of false-negative and false-positive signals.

There is a related issue that this thesis does not attempt to address. It is, in terms of the example in Table 3.16, whether if each line were subjected to an outcome-value (the auditor's overall materiality) of £1.075m, would the overall opinion be assessed to be within £1.075m? For example in the example described in Table 3.16, if each area was subjected to an outcome-value of £1.075m, is the overall materiality £2.15m, or £1.075m or some value in between? This is a problem, which has not been resolved. The value must lie between the two values. We can illustrate this in the following argument. Suppose there two account areas, as in Table 3.16, and that $\hat{Y}_1$ and $\hat{Y}_2$ are the expectation values and Y_1 and Y_2 the associated book values. Suppose also that the outcome-values are C_1 and C_2, and that they are both satisfied. Then the outcome–rules will be for account areas 1 and 2 will be: $|\hat{Y}_1 - Y_1| \leq C_1$ and $|\hat{Y}_2 - Y_2| \leq C_2$. What is the combined outcome-value? If the rules were additive we would have

$$|\hat{Y}_1 - Y_1| + |\hat{Y}_2 - Y_2| \leq C_1 + C_2$$
$$|\hat{Y}_1 + \hat{Y}_2 - Y_1 - Y_2| \leq C_1 + C_2$$

If the rules were such that the outcome-value was the overall materiality: say $C_1 = C_2$, then we would have: $|\hat{Y}_1 + \hat{Y}_2 - Y_1 - Y_2| \leq C_1$. Thus the overall materiality implied by the outcome-rules is between C_1 and $C_1 + C_2$.

3.9 What is a significant difference?

For the purpose of this thesis we define the significant difference for a particular AP to be the outcome-value. It should be related to the materiality level set by the auditor for the audit of the financial statements. There is no set function used to assign materiality to the area being tested. That is, a function that relates the overall materiality to the outcome-value. The Auditing Standards acknowledge this: '*Materiality is not capable of general mathematical definition as it has both qualitative and quantitative aspects*' (SAS 220.1). However in order to ensure some rational connection between the overall materiality and the outcome-value some firms adopt a rule that is for guidance and interpretation by the auditor. Two examples exemplify the different rules adopted by

the audit firms. One major firm uses a step function to assign materiality to the area being tested. It is similar to that set out in Figure 3.17 below. They use the term 'tolerable difference' which is equivalent to the 'outcome-value' used in this thesis. The table's use is best demonstrated by an example. Consider an account where the materiality for the financial statements was set against total expenditure. Suppose the value to be tested by an AP was £60m in a set of financial statements whose total value was £ 200m. Suppose also that the materiality was set at £2m. The amount being audited using the AP, represents 30% of the total expenditure. This lies in the band 25% to 50% in figure 3.17, and so for this example the outcome-value (or tolerable difference) for this test would be 0.33 of the overall materiality: that is £0.66m.

Value being tested by the AP as a proportion of total value of accounts.	**Tolerable Difference (outcome-value) as proportion of overall materiality.**
Greater than 75%	1.00
Between 50% and 75%	0.55
Between 25% and 50%	0.33
Less than 25%	0.25

Figure 3.17: Example of table for allocating materiality to account areas or balances. [Source: Adapted from the Audit Manual of KPMG].

In a second example, another organisation, the NAO, uses an empirical definition to assign materiality to the area being tested. The rule is:

$$\text{Outcome-value} = \text{Overall Materiality} \times \sqrt{\text{[Relative Value of the figure being tested to overall accounts]}}$$

In this example, the outcome-value is $\sqrt{[0.30]}$ of the overall materiality, which is 0.55 of the overall materiality: that is £1.1m. A justification given for this choice of function is that it is analogous to the theory for combining the variances for samples.

These levels for outcome-value: '£0.66m and £1.1m' represent quite different choices when it comes to judging the achievement of the AP. These examples are derived from a survey described in Chapter 7. While neither has a theoretical basis, they capture the need to allocate planning materiality across the different audit tests. Figure 3.18 below illustrates the different outcome-values produced by the two rules discussed above. The straight lines represent the first rule and the curve the second rule.

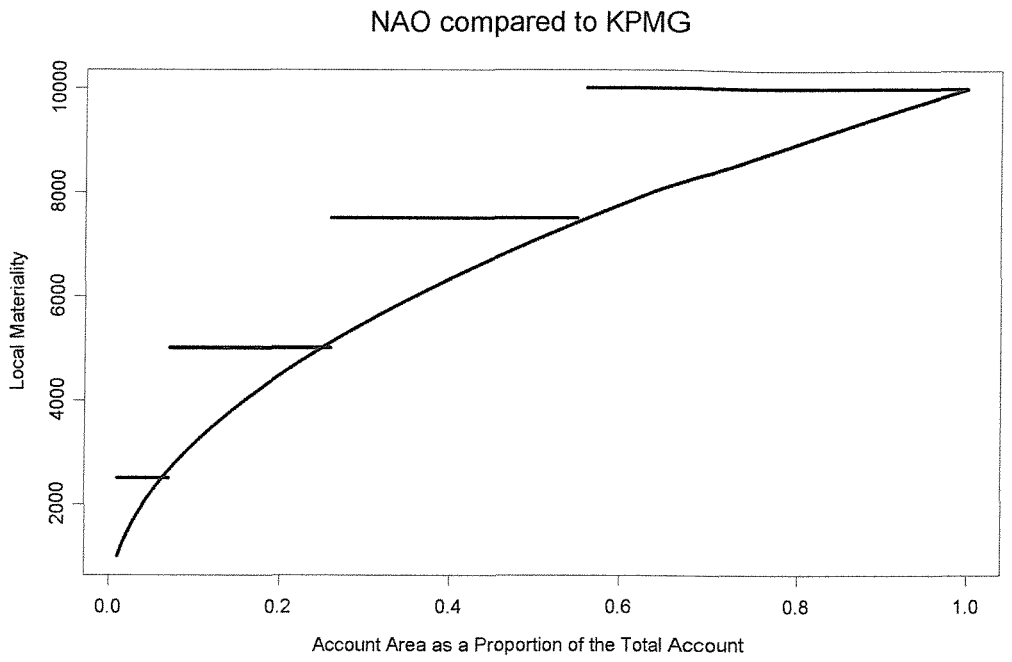


Figure 3.18: This graph shows the comparison of the NAO and KPMG methods of associating the outcome-values to overall materiality. The NAO rule produces the curved line while the KPMG model produces the series of horizontal lines. It will be seen that the NAO outcome-values are generally smaller than (tighter) or equal to the KPMG figure. [Source: Adapted from Audit Manuals of KPMG and NAO].

We will consider the subject of outcome-values further when we consider the results of experiments to see if particular APs reliably signal the presence or absence of error.

3.10 Seeking explanations for a significant difference

Once a significant difference is identified, the next step is to seek an explanation for it. The difference could arise from an incomplete model for the AP, which has not identified all the factors that make up the figure in the financial statements. Or it could arise from a potential error or misstatement. When developing the expectation values, and before the outcome of the AP is known, it is essential to identify alternative explanations for any potential difference between the value being audited and the expectation value. These then form the basis for making a judgement about the results of the AP, including any explanation provided by the audited organisation.

Why is this so important? The Joint Monitoring Unit of the ICAEW has identified this as a weakness of APs. Discussions with the JMU, revealed that part of the training of

the inspectors focuses on potential weakness that the JMU teams have identified in the application of APs. One such lapse cited in the training material is that '*no effort is made at an independent fair comparison of the expectation to the value being audited*'. The JMU also found that: '*the management explanations are too easily accepted when there should be a greater degree of professional scepticism*'. Libby (1981) and Anderson and Sechler (1986) have both addressed this issue and we will return to this in the review of the literature.

3.11 Conclusion

This chapter has described what an AP is and provides an overview of the types of AP used by auditors. It has presented three classifications of APs, those of the APB (comparison and relationships), the classification widely used in the academic literature (scanning, trend analysis, ratio analysis and modelling) and lastly that adopted in this thesis: namely single variable, two variable and multivariate procedures. The mapping between these classifications is in Table 3.13 on page 56. The classifications are consistent with one another. The grouping serves two purposes. First, it provides a convenient basis for comparison of what are vast ranges of procedures. Secondly, it provides a structure to assist any description of the complexity and potential reliability of APs. In the next few chapters we will use the classification of AP that is sensible for the content of the chapter.

Before the new work for this thesis is reported in Chapters 6 and 8, the Literature Review will consider the extent of use of APs and the perceptions of auditors of their use of APs. It will also look at experiments into the effectiveness of APs. The next chapter will commence that Review by looking at the requirements of those that set the Standards and some of the misgivings that they have expressed.

Annex A: An example: Pay Bill for a Government Agency

This example is based on payroll data from a government agency. The scenario is that the audit of this Agency and its paybill has been ongoing for many years and the current audit is for the financial year 1998-1999. For Government Agencies the financial year runs from April to the following March. Audited payroll data are available for the three years prior to the current year.

A.1: The data

There are data on two grade groups, higher and junior grades. The monthly details for the staff come from personnel. These data are not compiled for, or from, any of the financial accounting systems, and so is 'independent' of the pay and accounting systems. The system that produces the data on staff numbers has been reviewed, and found to be very reliable. Therefore there are no doubts about the reliability of the data.

The data for the example are given in Table 3A.1. The 36 months are data from the period before the current audit. The columns headed Pay show the monthly pay-bill in pounds. The other columns show the number of staff in post, by grade group.

	Monthly Staff No				Monthly Staff No		
Month	Higher	Junior	Pay (£)	Month	Higher	Junior	Pay (£)
1	244.5	567.0	711,467	19	252.5	573.0	846,313
2	242.5	572.5	726,544	20	255.5	577.0	831,168
3	244.0	570.5	723,623	21	255.0	577.0	830,465
4	244.0	570.5	716,063	22	255.0	577.0	828,350
5	245.5	575.0	724,356	23	254.0	574.0	830,647
6	249.5	579.0	720,067	24	256.0	568.5	837,838
7	246.5	579.0	757,157	25	256.0	562.0	843,286
8	249.5	577.0	760,633	26	256.5	550.0	906,033
9	251.0	580.0	750,155	27	256.5	555.5	883,164
10	253.0	573.0	744,563	28	256.5	561.0	1,011,073
11	251.0	573.5	747,895	29	258.5	560.0	953,292
12	251.0	573.5	758,952	30	261.5	560.5	932,600
13	251.0	562.5	768,286	31	263.5	560.5	931,937
14	258.5	563.0	823,000	32	267.5	557.0	938,760
15	256.5	567.5	916,200	33	269.0	555.5	939,753
16	255.5	565.0	846,277	34	270.5	561.0	929,673
17	255.5	567.0	853,349	35	270.5	560.5	963,927
18	254.5	564.5	837,579	36	277.5	562.5	974,645

Table 3A.1: The data for the pay example with the pay and staff numbers on a monthly basis. Periods 1 - 36 are the base data for constructing any model. [Source: Data from a Government Agency].

To emphasise that the intention is to predict the pay bill for months 37 to 48, the relevant staff numbers and the recorded pay bill for those months are in Table 3A.2.

Month	Higher	Junior		Pay Bill
37	278.0	561.0		997,620
38	278.0	556.0		991,664
39	279.0	558.5		1,052,095
40	279.0	558.5		1,020,729
41	279.5	562.0		1,035,535
42	279.0	560.5		1,028,753
43	278.5	559.0		1,172,642
44	279.5	555.5		1,056,564
45	276.5	564.0		1,053,587
46	277.0	564.5		1,017,679
47	279.0	560.5		1,097,499
48	277.5	558.0		1,072,966

Table 3A.2: Periods 37-48 are the monthly data for the year being audited. The pay data are the values to be predicted and are shown with the grey shading. [Source: Data from a Government Agency].

The basic information on pay is shown graphically in figures 3A.3 and 3A.4. Figure 3A.3 shows the monthly pay-bill for months 1 - 36. It will be seen that there is a steady increase, but there are some payments well above the sequence.

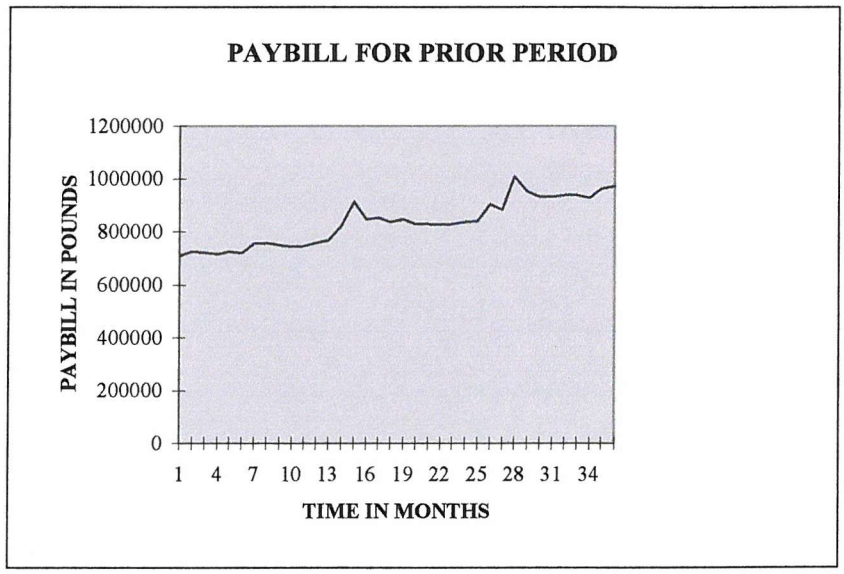


Figure 3A.3: Monthly pay bill for the three years prior to the audited year. [Source: Data from a Government Agency].

The auditor is aware that there are a number of features of the pay system that are relevant to the audit. First there is the annual cost of living rise for all grades. These pay rises are payable from the 1st April, but are not normally agreed until later. The pay rise for earlier months is then paid in the month following the agreement, and the pay rates amended accordingly. For the period of the example they are the same percentage for

all grades. Secondly, the pay-bill also includes advances of salary, which are also recovered through the system. These are assumed to be roughly in balance: that is the amount paid out each month equals recoveries. This assumption will be a factor in the final evaluation of the AP.

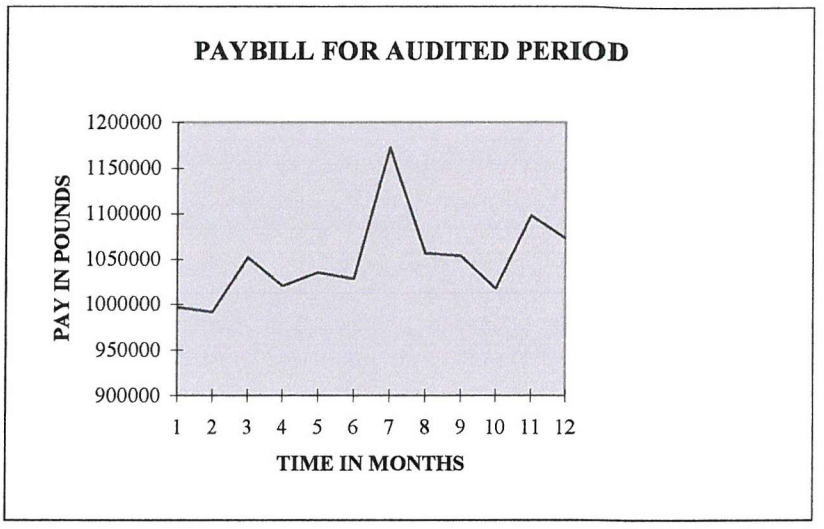


Figure 3A.4: Monthly pay bill for the audited year. Note that the y-axis starts at £900,000. [Source: Data from a Government Agency].

The staff numbers for the whole 48-month period are illustrated in Figure 3A.5. It shows that the staff numbers have been steady for the period, with minor fluctuations.

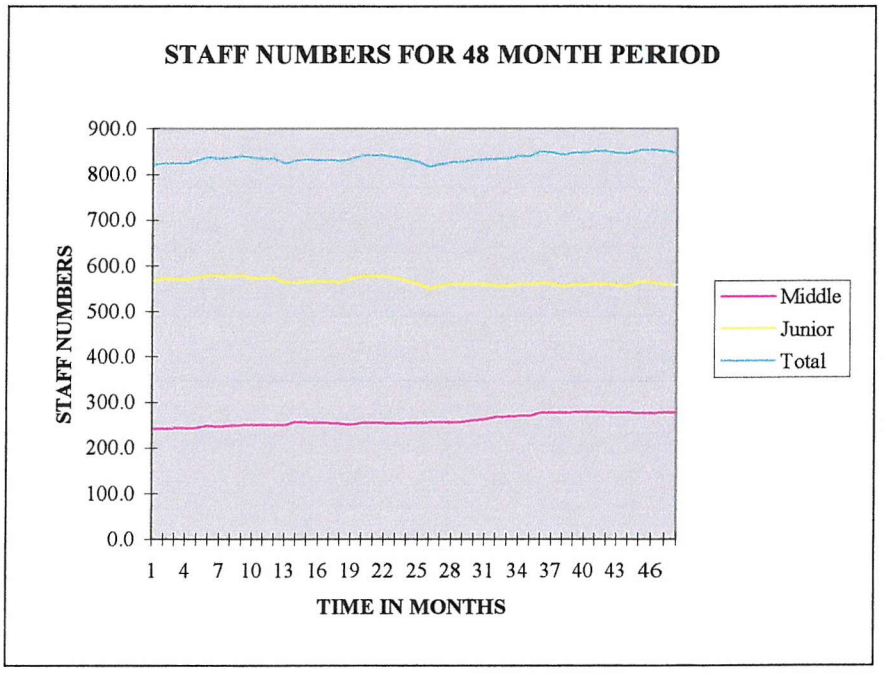


Figure 3A.5: Staff numbers for the whole period. [Source: Data from a Government Agency].

The annual charge to the accounts for pay is summarised in Table 3A.6. It shows the average staff complement by grade.

	Number of Staff		
Year	**Higher**	**Junior**	**Pay Bill (£)**
1995-1996	247.7	574.2	8,841,475
1996-1997	255.0	569.7	10,049,472
1997-1998	263.7	558.8	11,208,143
1998-1999	278.4	559.8	12,597,333

Table 3A.6: Average staff complements for each grade and total pay bill for each year.

A.2: Pay rise details for the period covered by the AP data

To enable the expectation to be based on a realistic relationship between the pay and staff data, allowance needs to be made for certain 'anomalies' in the data. Because of annual pay rises, the pay bill data is not on the same basis for all years or months. Two significant factors flow from these annual rises: back pay and the effect of inflation.

[a] Back pay

Back pay is significant because, for this organisation, the annual pay award is awarded from the first month of the financial year, but frequently the pay award is not agreed until several months into the year. The detail is set out in Table 3A.7.

Month	Pay (£)	Notes	Month	Pay (£)	Notes
1	711467	Start of year one data	25	843286	Start of year three data.
2	726544		26	906033	
3	723623		27	883164	
4	716063		28	1011073	Pay rise of 4.01% back-dated to month 25.
5	724356		29	953292	
6	720067		30	932600	
7	757157		31	931937	
8	760633		32	938760	
9	750155		33	939753	
10	744563		34	929673	
11	747895		35	963927	
12	758952		36	974645	
13	768286	Start of year two data	37	997620	Start of audited year.
14	823000		38	991664	
15	916200	Pay rise of 5.11% back-dated to month 13	39	1052095	
16	846277		40	1020729	
17	853349		41	1035535	
18	837579		42	1028753	
19	846313		43	1172642	Pay rise of 3.18% back-dated to month 37.
20	831168		44	1056564	
21	830465		45	1053587	
22	828350		46	1017679	
23	830647		47	1097499	
24	837838		48	1072966	

Table 3A.7: Data on the monthly pay, the dates of pay rises and the periods of backdating. The data from the audited year is shaded grey. [Source: Data from a Government Agency].

In this example, the pay award is due from months one, thirteen, twenty-five and thirty-seven, but the awards are paid in months fifteen, twenty-eight and forty-three. The payments in those months contain the pay rise for that month and the previous months for which it was due. For example the rise in year two, paid in month fifteen, included the back pay for months thirteen and fourteen. An example of the adjustment is as follows. Suppose the back pay is paid in month [j] and relates to months [j-1] and [j-2]. Calculate the element relating to the pay rise and add the relevant share to months [j-1] and [j-2]. The results of carrying out these calculations are shown in Figure 3A.8.

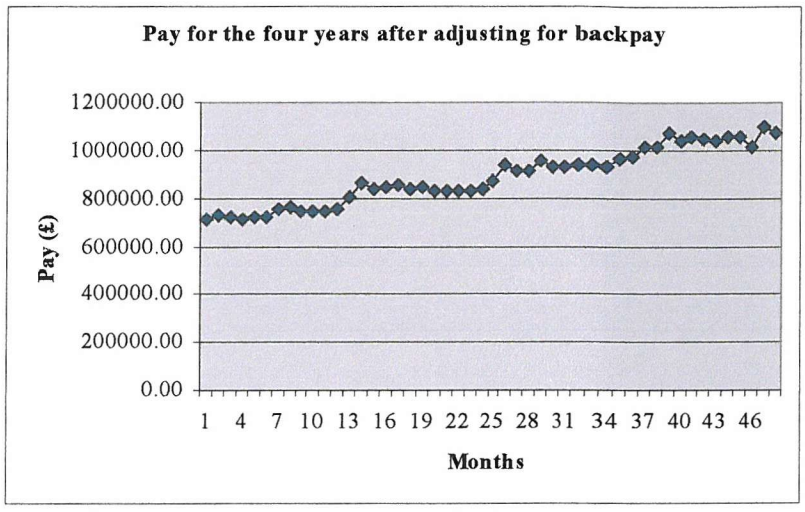


Figure: 3A.8: This graph shows the data after the distortions in the monthly pay amounts caused by the back pay have been removed. [Source: Data from a Government Agency]

Comparing the graph in Figure 3A.8 to those in Figures 3A.3 and 3A.4 we see that the large peaks have been removed. The next step is get all the financial data so that it is directly comparable to that of the audited year

[b] The effect of inflation

Inflation flows directly from the annual pay awards. If the rise in year two were 5.11%, then the pay in year one should be inflated by 5.11% to make the two years comparable. The same logic extends to all years. The data being audited is in year four; so all years should be adjusted accordingly. The pay rises relevant to this example are listed in Table 3A.9.

Year	Affecting Years	Rise
Two	1 - 2	5.11%
Three	2 - 3	4.01%
Four	3- 4	3.18%

Table 3A.9: A summary of the pay raises of the two prior years and the audited year – year 4. [Source: Data from a Government Agency].

We apply this information to the data so that it is on the same basis as year four. The data from months 1 to 12 is up-rated by 5.11%, then from months 1 to 24 is up-rated by 4.01% and finally from months 1 to 36 is up-rated by 3.18%. All the data is now in the same relative values as months 37 to 48: the data to be audited. The new data is in Table 3A10.

Months prior to audited period				Audited period	
Month	**Pay (£)**	**Month**	**Pay (£)**	**Month**	**Pay (£)**
1	802545.04	23	891429.74	37	1019273.71
2	819552.12	24	899146.95	38	1013317.71
3	816257.19	17	915792.97	39	1073748.71
4	807729.40	18	898868.99	40	1042382.71
5	817084.02	19	908242.11	41	1057188.71
6	812245.97	20	891988.87	42	1050406.71
7	854084.02	21	891234.43	43	1042719.71
8	858005.00	22	888964.66	44	1056564.00
9	846185.67	23	891429.74	45	1053587.00
10	839877.81	24	899146.95	46	1017679.00
11	843636.35	25	904559.11	47	1097499.00
12	856108.81	26	969301.47	48	1072966.00
13	867631.73	27	945705.23		
14	926349.43	28	939855.27		
15	896990.58	29	983606.69		
16	908203.47	30	962256.68		
17	915792.97	31	961572.60		
18	898868.99	32	968612.57		
19	908242.11	33	969637.15		
20	891988.87	34	959236.60		
21	891234.43	35	994579.88		
22	888964.66	36	1005638.71		

Table 3A.10: The pay data after adjustment for back pay and the effects of pay inflation. The pay relating to the audited year is shaded in grey. [Source: Data from a Government Agency].

A3. Calculating the expectation value.

Now the data has been adjusted for the back pay phenomenon and the effect of pay inflation, the next step is to build the model to estimate the expectation value. This is calculated using least square linear regression. Using the staff data from Table 3A.1 and the pay data from Table 3A.10 we can estimate the model. The dependent variable Y is the monthly pay bill, while the higher grade staffs are explanatory variable 1 (V_1)

and junior staff are variable (V_2). The estimated parameters are 913680.50 (intercept), 7239.80 (variable one) and -3394.90 9variable two respectively. The model is thus: $Y = 913680.5 + 7239.8 \times V_1 - 3394.9 \times V_2$. All figures one decimal place. The model developed is then applied to the staff data for the audited period to produce twelve expectation values, one for each month. These are combined to produce the expectation value for the year. The results are in Figure 3A.11.

Account Value	Expectation Value	Difference	Difference as %
£12,597,333	£12,341,780	£255,553	2.09%

Figure 3A.11: The results of regression based AP on payroll. [Source: Data from a Government Agency].

Annex B: Examples of Frauds that an AP would have discovered.

Laribee

This organisation had stock at many different locations. There was a variety of reliable information available. An AP that compared stock levels with the maximum capacity levels by location of the stock would have indicated a potential problem in the existence of stocks. By using non-financial data (capacity) that was reliable and could be measured by the auditor, a reasonableness type AP would have directed the auditor's attention to an overstatement of stock in the accounts.

Mattel Inc.,

Mattel is an organisation that has many competitors in the same markets. An AP on disaggregated data (monthly), relative to the competitors in the same field of business, would have indicated unusual fluctuations in the sales. Using annual data would not have revealed the problem. Also a reasonableness analysis using non-financial data (specific contract information) would have identified an under-valuation of royalty expenses.

Regina

Financial statement comparisons over a three-year period would have shown that accounts receivable and inventory had both increased by significant percentages between 1986 and 1988. Collectively they made up 76% of the assets in 1988. Activity ratios would have reinforced the concern and placed the auditor on enquiry.

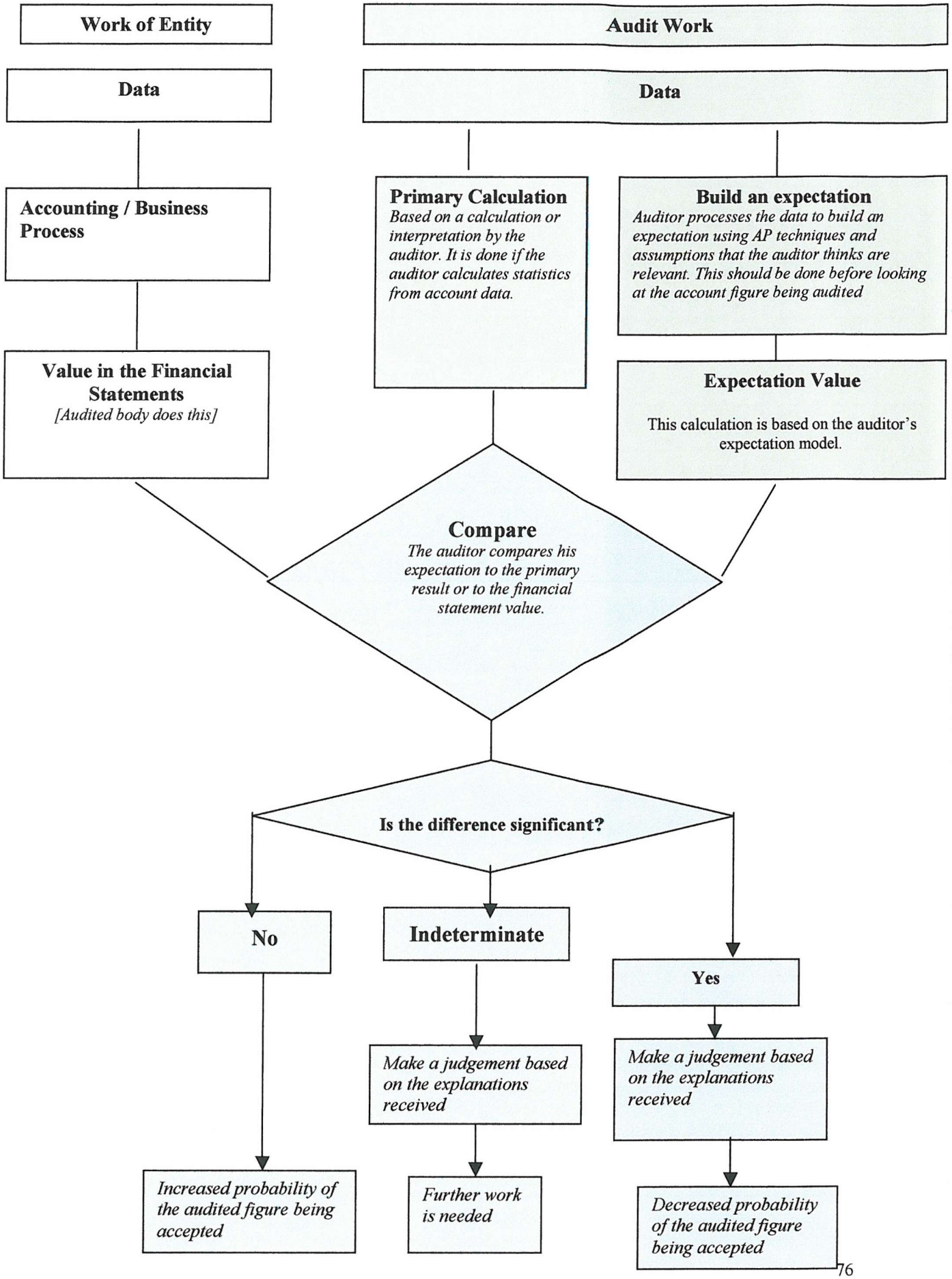
US Surgical Corp

A thorough analysis of the related accounts within the financial statements would have alerted the auditor to potential problems. Trend analysis performed at the planning phase, on all balance sheet and income-statement accounts, together with an understanding of the relations between the different accounts, would have identified fluctuations, that proved to be material in what are typically low risk account areas. For example, the inventory nearly doubled as a percentage of total assets and inventory turnover declined.

Zzzz Best

There were many companies in the business that Zzzz Best dealt in. Zzzz Best's restoration contracts were over 200 times the average for the industry. A simple AP comparing restoration revenue trends with those in the industry would have indicated potential fictitious revenue.

Annex C: Model of the Analytical Process.



Chapter 4: Review of the Literature on APs

4.1 Introduction

The objectives of the literature survey include gathering information as to:

- why APs are used,
- what techniques are available and how they are used,
- when they are used, and
- how auditors perceive APs.

This information, as well as being interesting in its own right, can help an assessment of whether APs are effective, economic and efficient. In other words, are APs useful in the sense that they are an inexpensive tool-set (economy criterion) and that they meet the required standards in terms of accuracy and achieved assurance (efficiency criterion). Also, is their application capable of producing a result that has a low probability of giving false signals about the presence or absence of material error (effectiveness criterion)?

The review of the literature will be considered in this chapter and Chapters 5 and 7. After a brief historical overview of APs, this Chapter is principally concerned with the standards set by the APB, Parliament through the Department of Trade and Industry (DTI) and the legal courts. Critical to the implementation and maintenance of standards is the integrity and professionalism of auditors, so that will also be considered. We then look at a number of surveys that have been published in the academic literature since 1980. This will be done in Chapter 5, which addresses the questions of what kinds of APs are used, and when are they used during the audit cycle. The question of why APs are used will be addressed by considering auditor's perceptions of APs. Chapter 6 will introduce a new piece of work, a survey of the policy makers within the largest audit firms. It will address some of the questions from the literature review. The technical partners are responsible for the promulgation audit policy within their audit firm. They may have a different perspective from the practitioners within the firms. Thus, Chapter 6 will confirm and contrast the earlier surveys reported in Chapter 5.

In Chapter 7, the perceptions of audit practitioners about APs will be contrasted with the results of academic research into the effectiveness of APs. There have been many

investigations into the effectiveness of APs in reducing other forms of testing. The work will then be supplemented by the results of some new experiments to test the effectiveness of APs in detecting errors. These will be reported in Chapter 8.

The literature review concentrates on the substantive stage of audit. This is, in part, because most of the research into effectiveness of APs concentrates on this phase. Another reason is found in the rationale for using APs. At the substantive stage their usual purpose is to provide evidence to confirm the book value in the financial statements. In contrast at the planning stage their purpose is to direct attention, to highlight areas that might require further attention. It is not necessarily that there is something wrong, but merely that there could be. Another signal that an AP may give at the planning stage is that no further attention is necessary. If an AP, at the planning stage, is used to justify no further work, then effectively it has become substantive: it is the auditor's only evidence of that audited value and the Standards require that some substantive testing should be done. So the AP used at planning now becomes a substantive AP. By considering APs at the substantive stage, one can extend the argument to planning APs. If it should transpire that APs are not used at the planning stage then that provides further justification.

4.2 Brief Overview

As described in Chapter 3, wide ranges of techniques are used as APs. For many years the collective noun for such methods was Analytical Review, and this is the term used to describe APs in some of the literature. Stringer (1975) attributes the first official use of the term Analytical Review to SAS Number 1 issued by the AICPA in 1972. But APs had been in use for many years before that. Specific APs are described in the Audit Manual of Coopers and Lybrand issued in 1969: examples included the use of gross profit as a percentage of turnover of stock, and the ratio of year-end stock to turnover. Stringer (1975) reported that APs had been in use by his firm, Haskins and Sells, for over forty years! Until Stringer's (1975) article there had been little reported academic work. It appears that for a long period of time, many auditors used APs such as comparisons, or made predictions based on their data, but had no collective name for APs. Thus Lev's (1980) comment is appropriate: '*In fact the whole subject of analytical review, appears to be shrouded in mystery. Accounting firms have allegedly*

been conducting analytical reviews for some time. Yet judging from the literature, no established widely known procedures have emerged from this long practice'.

The year after Stringer's (1975) article, Albrecht and McKeown (1976) documented the use in the early 1970's, of ratio, trend analysis, regression analysis and time series analysis (univariate and bivariate methods). Eleven years later, Loebbecke and Steinbart (1987) reported the use of very simple techniques such as scanning, simple trend analysis and simple ratio analysis as substantive evidence sources, most applications using annual, not disaggregated, data. At least one firm is still a major user of statistically based APs: since the early 1970's Deloitte Touche (or 'Deloitte Haskins and Sells' as they were then called) have applied regression analysis as part of their audit strategy. Stringer (1975), along with Stewart (1978), was responsible for much of the development work. They reported that regression based APs were used in up to ninety percent of the firm's audit engagements in the USA. The typical application used a base of thirty-six observations in order to make a prediction for the twelve months of the year being audited. It was applied in the audit of pay, production costs, and travel expenses and interest paid and received. By 1978, Deloitte Touche was applying regression more than 30,000 times a year, using a package called STAR[1]. Deloitte Touche still uses the regression based APs; STAR has been updated several times in that period.

4.3 The Integrity and Competence of Auditors

APs involve value judgements. Thus the competence, integrity and independence of the auditor are vital if APs are to provide the reliable, audit evidence that is necessary to support the audit opinion on a set of financial statements. Turley and Cooper (1990) found that there was unease amongst practitioners about APs. More damning was the finding that there was little understanding of the theory underlying many audit techniques: *'Interviews with audit staff throughout the profession revealed fairly widespread divergence between theoretical definition and practical understanding.'*

Auditors, in their reluctance to become too involved with theory, do not concern themselves with these issues. Professor David Flint, emeritus Professor of Accounting at the University of Glasgow, succinctly summed up the situation. He wrote: *'The*

[1] **STAR** - Statistical Techniques for Analytical Review in Auditing.

practice of Auditing has certainly developed without the prior formulation of any theory; and the public conception of the audit phenomenon has evolved without the assistance of a theory'. He went on to say: *'that it is represented as a 'practical' subject, with the inference that auditing is what auditors do, and that it does not have a theoretical content'* (Flint (1988)). There still appears to be some truth in this statement, at least in the eyes of practioners. David York, the national technical director of audit for the firm Moores Rowland, sums up the rationale of many auditors. He wrote: *'The underlying reason that the concept of separate assessment of IR and CR has some degree of acceptance in a mistaken belief that the audit risk model is a formula that can be used to decide 'how many to do'. The sooner it is returned to the box labelled 'theory - dangerous if taken too seriously' the better'* (York (1997)). While York has a valid point about the Audit Risk Model being imperfect, the reaction is flawed. It should not be that the concept of theory is wrong, but that the particular piece of theory has its defects. Keynes (1936) distinguished two extremes: *'Practical men, who believe themselves to be quite exempt from any intellectual influences, are usually the slaves of some defunct economist. Madmen in authority, who hear voices in the air, are distilling their frenzy from some academic scriber of a few years back'*. It may be that auditors consider themselves to be firmly in the former category.

This view of the auditor's suspicion of a theoretical approach has support from the research of other academics. Such research has also raised doubts about the ethics, independence and competence of auditors. The work of Humphrey and Moizer (1990) is a good example. As part of their research into the audit function, they interviewed 18 experienced audit managers. They found that there was a growing marketing orientation of auditor's work, the selling of services to the client being a major aim of an audit manager. In addition there was a perception amongst the managers that fees gave the audited body too much influence over the auditors. A typical comment from a manager was: *'As the market place becomes more competitive, we have got to do cheaper audits, do them quicker. This means continually reassessing areas where we need to put time into, where we might have commercial exposure, and reduce time in area where we don't have commercial exposure'* (Humphrey and Moizer (1990) pp229 - 230).

The quest to improve efficiency is a natural one. In order for it to be done in a way that can be defended to the world outside of the audit profession, improvement must be directed, and implemented, by auditors who understand the interplay between theory and practice. Auditors, if they are going to make judgements about a procedure, or the results of applying it, should understand the principles on which it is based.

Humphrey and Moizer found there was an inability on the part of the auditors to explain the logic behind procedures that were applied as part of the audit. At the time of the survey, sampling was a major audit procedure, and so it was the example used in the interviews. *'The audit managers were asked to explain the relationship between the sampling plan and the overall level of confidence to which they worked when arriving at their audit opinion. Their replies revealed that for most of them it was an act of faith'.*

It was also found that there was an unwillingness to follow the Auditing Guidelines (replaced by the Auditing Standards in 1995) laid down by their own professional body. The Guidelines were viewed as general and high level: *'setting a minimum standard for the smaller firms'* (in the words of one manager quoted by Humphrey and Moizer). In fact that perception is correct in that, the Guidelines and the Current Standards are the minimum standards that all auditors are required to adhere to. Worryingly the managers involved in the survey felt that the use of such guidance was: '*as a legitimising device, used to justify the need for a particular audit test to be performed in any external dispute*' with the audited body. As a result of the survey Humphrey and Moizer concluded that auditors lacked independence. They put it thus: *'In documenting the changing nature of audit practice and the increasing emphasis on risk-adaptive, business-orientated approaches, one particular issue stands out – namely, that of auditor independence, or to be more precise the portrayal by audit managers of an audit function increasingly dependent on the dictates of corporate management'.*

Because of their work, Humphrey and Moizer questioned the auditor's competence and integrity. They concluded that: *'Indeed, any manager concerned about her/his professional independence could scarcely have described so freely the growing business and marketing orientation of their work and their inability to explain the logic behind sampling procedures and choices as to the extent of their audit work. Ultimately*

it is this contrast between the auditors' faith in their professional integrity and competence, and the scepticism and concern of interested observers, user groups and regularity organisations that is central to the notion of an audit expectation gap'. They are not alone in articulating this concern about the 'expectation gap'.

The independence, competence and integrity of auditors is particularly significant for APs, which rely greatly on judgement, knowledge of the audited entity, expertise in the use of the procedures and a willingness to question sceptically management's assertions about figures in the financial statements.

4.4 Have the SASs had any effect?

The independence of auditors from the management of the audited entity and their integrity and competence are three attributes that are keystones in the maintenance of auditors' professional standing. One question is whether the SASs which came into effect for accounting periods ending on or after 23 December 1995 have had an impact on addressing the potential problem of auditor independence and integrity, in particular as they apply in the application of APs. For a view of this we consider the work of academics including Beattie, Brandt and Fearnley (1997) on auditor independence, and the review by the Joint Monitoring Unit (JMU) of audit working papers (integrity and competence of auditors).

Beattie, et al sum up the importance of the auditors' independence: '*Auditors' independence is fundamental to the value of the audit service and audit is an essential part of the regulatory framework that maintains the integrity of our capital markets*'. This is too narrow; such concepts do not apply solely to the financial markets. It is also crucial for good government both at the national and local levels. In the public sector, audit is an essential part of open, democratic, government. For example, independence is enshrined in the statutes of the Supreme State Audit Institution in the UK (the NAO), and the Public Accounts Committee – the senior committee of MPs in the British Parliament guard this jealously. The Head of the NAO, the Comptroller and Auditor General, Sir John Bourn is an officer of the House of Commons.

Beattie et al go on to define independence as '*acting with integrity and objectivity and being able to withstand pressure from management to infringe professional standards*'. This captures the importance of audit integrity and objectivity, both of which are key to

the successful application of APs. It is necessary to understand what pressures the auditor is under to provide a clear certificate for an account, be they real or perceived pressures. This is important since APs are a key part of the audit program; Ameen and Strawser (1994) found that over 40% of audit time was given over to the use of APs.

Beattie et al claimed to identify factors that are perceived to be key in enhancing audit independence and several factors that threaten that independence. This information was obtained from a survey of Finance Directors and Audit Partners with 153 (51%) and 244 (80%) responses respectively. They asked the two groups to choose, from lists provided, the most important regulatory or economic factors that either enhanced or undermined auditor independence. Although the two groups identified different factors that were perceived to enhance or threaten the independence of auditors, they identified common themes for these factors.

Broadly the factors that are perceived to enhance auditor independence are regulatory. Examples included the risk of loss of registered auditor status, and risk of referral to the Review Panel. On the other hand the factors that undermine auditor independence were economic. Examples included an audit firm being dependent on one client for more than 10% of the fee income or audit fee discounting.

The paper by Beattie et al reinforces the theme of Humphrey and Moizer (1990): there remain worries about 'auditor independence', even since the new Auditing Standards were issued in 1995.

4.5 The JMU

While Beattie et al provide an insight into perceptions of the factors that threaten or enhance the auditor's independence, a related issue is the extent to which auditors comply with the Standards that guide the profession. In this connection, we can consider the views of the regulator of the audit profession, the JMU, whose role is to monitor compliance by the audit firms with the audit regulations and to assist in the raising of standards. One of the ways JMU does this is by a series of visits to the firms to review the audit working papers.

In an article for the journal of the ICAEW, Accountancy, Garvey and Dietz [2] (1997) summarise some key issues relating to APs that arose from the JMU's work. Although this was only two years after the introduction of the standards, all audits completed after the 23 December 1995 had to comply with the Standards. Since the profession was extensively consulted in the development of the Standards there is little excuse for not implementing the Standards. This is important, because the JMU article is well after the date when all Standards were mandatory and is therefore a useful indication of how the profession complies with the Auditing Standards and the auditor's integrity some seven years after Humphrey and Moizer (1990).

Garvey and Dietz report that the JMU found that the working papers did not document the consideration of the threats to the audited entity. As audits develop new facts emerge that may require audit firms to revise their records and their audit approach. Such information is basic to ensuring that a proper audit can be conducted. The JMU found that: *'The planning and permanent papers do not provide this information.'* Decisions were not explained, a particularly important attribute in a process that requires expertise and judgement. For example: *'Too often the working papers lack evidence that auditors had considered the possible effects of relevant laws and regulations, or that the potential risks of fraud or error, have been considered'*. They go on to say: *'It is not sufficient to tick a box - there needs to be positive affirmation even where no matters require attention'*. In a comment about the updating of information about the audited entity, the JMU concluded that it: *'unfortunately is often poorly documented'*. The poor justification of audit decisions was common. *'SAS230 includes two standards that stress the need to record audit procedures, matters of judgement and conclusions thereon. Recording is probably the single largest area of regular concern on visits. SAS 230 needs more than just a tick in the box on an audit programme, and if a judgmental matter arises this should be explained'*. Garvey and Dietz also concluded that there was no evidence that the auditors had understood the business, a key requirement stipulated in SAS 210 on Knowledge of the Business and Accounting and in SAS 300 on Internal Control Systems and Audit Risk Assessments.

On the subject of APs the JMU commented that: *'Firms should consider what they would expect the client's results to be, rather than religiously comparing this years*

[2] Hazel Garvey is a technical and training manager and Alan Dietz is an inspector with the JMU.

recorded results to last year's'. The emphasis is that of the authors. Here we have evidence that APs are applied without the development of expectation values. This reference was to substantive APs, since the JMU found that APs were not used at planning or review; *'APs should be applied at both the planning and overall review stages, but this rarely happens'*. This provides evidence that the standards are not being followed: the application of APs at these stages of the audit is a standard with which auditors are required to comply.

Also of concern was the failure to connect materiality to the extent of audit testing: Materiality and the Audit (SAS 220) and Audit Sampling (SAS 430) being the relevant SASs. *'The responsible individual should use his or her professional judgement to assess audit risk and to design/agree audit procedures to reduce risk to an acceptable level. Materiality is generally being computed, but then not used, especially when fixing sample sizes. In many instances, sample size and selection are not explained and/or fail to address material items, and the samples taken are not representative of the whole population'* Garvey and Dietz (1997). The JMU also found that *'some firms are still not following up on all items selected. In respect of those circularised about debtors, about 60% reply - but about 40% do not'*! This criticism goes to the heart of the concern that auditors do not understand the procedures they employ and do not follow through on the audit work. Although this relates to sampling, and not APs, it demonstrates the auditor's failure to connect materiality to the amount of substantive work they should do. The JMU found instances of error being dismissed purely on the grounds of not being material. There was no questioning of *'how did it arise, and could a material error have arisen from the same cause?'*(Garvey and Dietz (1997)).

The JMU were also worried about management representation. The inspectors commented: *'It is not enough merely to accept a written representation from the directors'* Garvey and Dietz (1997). References to this were made when considering the SASs relating to 'Going Concern' (SAS 130) and 'Subsequent Events' (SAS 150). These are usually considered by the audit senior/manager at the end of the fieldwork. The implications for APs are serious, when seeking explanations for significant fluctuations identified by the APs, the auditor has to obtain management responses. Being too willing to accept, without proper questioning, the responses will undermine the AP.

The work of the JMU confirms the concerns raised by Humphrey and Moizer (1990). This indicates that professional competence and integrity were, in 1997, still 'not proven'. For example, the failures to understand the business and to develop expectation-values are a particular concern for the application of APs. The requirement for both is neatly summarised by Knechel: *'Analytical review procedures rely upon an auditor's ability to generate an expectation of an account based upon knowledge of a company's business and operating environment* ' Knechel (1988a).

Unfortunately the JMU appears to have been less forthright since the paper of Garvey and Dietz (1997).

4.6 The Legal Setting and DTI Reports

Individuals and organisations are challenging the audit profession and seeking redress for what they consider the failings of auditors. Where parties try to address their grievances through legal avenues, many claims are settled out of court for fear of creating a legal precedent. An example is Peat Marwick who, in August 1991, settled a claim out of court and paid out over forty million pounds. Following the failure of the Maxwell organisation, there was an out of court settlement by the auditors (Coopers and Lybrand) of £68m to the liquidators and a further undisclosed sum to the defrauded pension funds[3]. Such was the impact of the failure of the Maxwell financial empire that there was a DTI Report into the failure, the scope of which included the role of the auditor.

Other sanctions for incompetent or inadequate audit work occur through the Joint Disciplinary Scheme. For example, the Joint Disciplinary Scheme tribunal imposed fines on Coopers and Lybrand totalling £3.5m for inadequate work in the audit of the financial statements of the Maxwell organisation. This should not be regarded as an isolated incident. For example, the May 2001 issue of Accountancy reported that Coopers and Lybrand had been fined £100,000 by the Joint Disciplinary Scheme tribunal. This was because of its role as auditor to Resort Hotels, and related to a fraud that should have been detected by the auditor.

[3] May 2001 Accountancy, page 9

4.7 The Courts

Whether APs will be defensible in a court of law, is an important issue from the perspective of this thesis. This is now examined from the viewpoint of how the courts have considered procedures similar to APs. Indeed will the courts rely on the auditor as an expert and not probe the techniques, assumptions and judgements that form part of the AP process? How will they react to a statement such as that by Dacoca and Holder (1985): *'APs were used merely because they had been utilised in the past without consideration of their diagnostic ability'*. Like Dacoca and Holder, Turley and Cooper (1990), Humphrey and Moizer (1990)) and Garvey and Dietz (1997) all questioned the auditors' understanding of the procedures they use.

This section looks at court cases that relate indirectly to APs, there being no cases where an application of an AP has been challenged directly. Although some of the court cases cited are from the USA and therefore set no explicit precedent for the UK, they are from legal systems that have a common heritage with the UK and so provide some light on legal thinking.

In the USA, Wallace (1983a, 1983b and 1983c) investigated a subset of APs, namely statistically based techniques such as regression techniques. Wallace was unable to find examples of regression based APs that had been considered by the courts. So she examined legal cases from other sources where the use of regression had been a factor in the court's judgement. Wallace surmised that the techniques would have to be clearly explained to the court. In one case, where the court found the graphical representation of simple regression (involving one predictor [independent] variable) to be 'persuasive', it rejected the use of multiple-regression. This was partly because the criteria used were subjective, but 'in the larger' part because the testimony of the expert witness was '*not comprehensible to the court*'. An inference is that the courts are likely to accept methods that they can comprehend.

That finding has been confirmed by a judgement[4] of the Court of Appeal, Criminal Division in 1996, which included the following statement: *'Evidence of the Bayes Theorem or any similar statistical method of analysis in a criminal trial plunged the jury into inappropriate and unnecessary realms of theory and complexity, deflecting*

[4]Reported in The Times, May 9, 1996

them from their proper task. Their Lordships ... had very grave doubts as to whether that evidence was properly admissible because it trespassed on an area peculiarly and exclusively within the jury's province, namely the way in which they evaluated the relationship between one piece of evidence and another. The Bayes Theorem might be an appropriate and useful tool for statisticians, but it was not appropriate for use in jury trials or as a means to assist the jury in its task.'

The relevance to APs is obvious; APs are about the relationships between sets of data (evidence). As in the US, the courts seem to be saying that the evidence must be clear and interpretable to a layperson and should not rely upon a theoretical concept. If the application of an AP were disputed in a court of law, would it mean that an auditor saying, 'my opinion is founded on my experience' has greater force than an opinion that is supported by evidence based methods? It is not clear whether that means that an easy-to-understand procedure such as simple comparison would be more acceptable than linear regression or some other modelling procedure that used, say, a computer program or complicated spreadsheet analysis.

Professor Adrian Smith commented on this unwillingness to accept an evidence-based approach in his inaugural address[5] as president of the Royal Statistical Society in 1996. *'So there we have it. To hell with rationality as we know it - their Lordships have pronounced! Perhaps as a consequence, the style of debate and enquiry, which characterises much of the working of the UK parliament and its committees stresses and rewards the mastery of an adversarial style, employed for short-term effect, rather than any long-term commitment to an evidence-based approach. The discussion of our national affairs therefore takes place in an overwhelmingly superficial and silly atmosphere, which runs counter to the more dispassionate forms of evidence gathering and assessment that should characterise a mature democracy'.*

While this comment was directed at Government, the implication for APs is that it might be preferable to have uncomplicated procedures that are not based on theory. There is a misunderstanding here. Simple methods can be rational, and theoretical (or evidence based) procedures can be simple. These are not exclusive conditions. The

[5] The Address of the President, delivered to The Royal Statistical Society, on Wednesday, June 12th, 1996

parsimonious principle should apply and the procedure/model developed should be as simple as possible using the smallest number of parameters consistent with describing the important features of the data. The AP used must be transparent, easy to understand, convincing and accepted by expert witnesses and juries.

4.7.1 A particular judgement: Pacific Acceptance v Forsyth and Ors (1970)

There is a particular legal judgement that has implications for the use of APs as a source of evidence. The case is described by Godsell (1991) and involved a company called Pacific Acceptance suing the auditors claiming it had lost three million Australian dollars as a consequence of the auditors' negligent failure to investigate and report certain irregularities. The auditors sought to rely on the defence that they had complied with professional standards, the errors that they discovered had emerged in the course of work that was additional to that required to comply with the standards, and that the management failed to advise them of irregularities of which they (the management) were aware. In fact the auditors had not appreciated the significance of the errors.

It was a comprehensive judgement concerning auditors' duties, and was the landmark Australian case of *Pacific Acceptance v Forsyth and* Ors (1970) 90 WN (NSW) 282. The judgement consists of an exhaustive evaluation of auditors' duties and responsibilities, and also covered the standards of skill and care, which should be exercised, in obtaining audit evidence. The main consequences for the auditor from this comprehensive judgement were summarised by Justice Moffit in terms of ten key principles, which are set out in Table 4.1.

There are seven principles that are particularly pertinent to APs and they are described below and summarised in Table 4.2 on page 94. The numbering refers to the judgement described in Godsell's book and is consistent with that book.

1.	Auditors have a primary duty to audit, which encompasses a continuous duty promptly to warn management of suspicions that fraud or error may exist, and which are not constrained by the duty to report.
2.	The duty to audit encompasses a duty to pay due regard to the possibility of fraud, and actively to investigate the possibility of fraud, in circumstances in which suspicions are, or should be, aroused.
3.	Auditors have no duty to detect fraud or error in the absence of circumstances which should arouse their suspicions, or merely because fraud or error exists, in the financial statements.
4.	It is the duty of auditors to obtain sufficient, relevant and reliable evidence to satisfy themselves of the various matters necessary to form their opinion, and they are not entitled to rely purely on management in this regard.
5.	Auditors are not to be excused from negligent conduct on the grounds that the client's directors or management were also negligent, or even if they were fraudulent.
6.	Reliance on independent sources of evidence is an aid and not a substitute for an auditor's procedures.
7.	Professional standards and practices must reflect changes in the economic and business environment and are not limited only to those adopted by the majority of the profession. Professional standards provide a guide, but the courts will exercise their prerogative to assess the reasonableness of professional standards in specific circumstances and therefore what constitutes reasonable care and skill.
8.	A qualified report must be clear and unambiguous and must, where possible, identify and quantify the matters of qualification.
9.	Before placing audit reliance on internal controls, these must be thoroughly ascertained, evaluated and tested.
10.	Auditors must carefully plan, and supervise all of their work, including work performed by subordinate staff.

Table 4.1: Ten key principles for auditors set out by Justice Moffit in the *Pacific Acceptance v Forsyth and Ors* (1970) case. [Source: Extracted from Godsell (1991)].

4.7.1.1 Error and Fraud are the auditor's concern

Justice Moffit addressed the issue of the auditor's obligation to discover error, fraud or other misstatements. The judge observed that: *'It is clear that in planning and carrying out his work an auditor must pay due regard to the possibility of error or fraud. Once it is accepted that the auditor's duty requires him to go behind the books and determine the true financial position of the company and so to examine the accord or otherwise of the financial position of the company, of the books and the balance sheet, it follows that the possible causes to the contrary, namely, error, fraud or unsound accounting, are the auditor's concern'* (my underlining). This can be interpreted to mean that the auditor has a duty to be aware that there could be error or fraud. As a consequence the

auditor should be using procedures that will detect error or fraud that is material. For an AP this means that if material error exists the AP must have a reasonable probability of detecting that error.

In the third principle the judge sought to balance the requirement for the auditor to have in place procedures that have a reasonable chance of revealing material fraud, by saying that auditors have no duty to detect fraud or error in the absence of circumstances which should arouse their suspicions. The judgement spelt this out at great length. A key passage was: *'An auditor pays due regard to the possibility of fraud or error by framing and carrying out his procedures, having in mind the general and particular possibilities that exist, to the intent that if a substantial or material error or fraud has crept into the affairs of the company he has a reasonable expectation that it will be revealed. The problem is an intensely practical one; on the other hand, it may be unjust to criticise a procedure, particularly with hindsight, merely because it was not apt to reveal some fraud devised with particular ingenuity, or some isolated or minor fraud or error. In such cases in particular it is important with resolution to exclude the operation of hindsight because after the event it is often easy to think of procedures that could have been adopted that would have revealed even the ingenious fraud, whereas in fact the auditor looking at the matter as it then presented was acting reasonably'.*

4.7.1.2 Management Representation

The judge expected that an auditor would obtain *'sufficient, relevant and reliable evidence'*, and should treat with scepticism any submissions from management. This is covered in points 4 and 5 of the judgement. Particularly relevant is: *'Having in mind the function of the auditor and his relation to the shareholders, and in particular his duty to them in relation to the directors, I do not find merit in a submission which in effect is that although the auditors were negligent they should be excused because the directors were also negligent'*. Thus in respect of the auditor's dealings with management, the auditor in choosing to rely on management representation must corroborate those and take ownership of the subsequent evidence. It is not a defence when something goes wrong in the audit to say that management misled me and so it is their fault. In the AP context, this can be interpreted to mean that when seeking explanations from management for variations from the auditor's expectation value it is the responsibility of the auditor to adopt a sceptical attitude to management's

explanations and not to see them as a justification for not pursuing the potential discrepancies.

4.7.1.3 The audit evidence must be the auditors'

The auditor must take responsibility for any evidence that is used to signal the presence or absence of material error. It is not a justification to use the defence that another party informed the auditor or provided the information. In relation to this point (point 6 of the judgement), the judge stated: *'Prima facie the auditor's job is to check material matters for himself from available documents and he does not ordinarily do his job or 'audit' if he merely seeks the assurance of another as to the check that other has made or as to his views as to the effect of documents'.* The evidence produced by other parties is an adjunct to and not a substitute for the auditors' own work. This taken together with point 4 implies that explanations sought from management must be considered with audit scepticism and not unintelligently accepted. This point also applies where an AP uses data that is from a source within the organisation that is outside of the financial systems or is external to the organisation being audited; the accuracy of such data should be checked.

4.7.1.4 Reliance on systems must be evaluated and tested

The audit of internal controls was a particular feature of the *Pacific Acceptance v Forsyth and Ors* case. The judge said that before placing audit reliance on internal controls, these must be thoroughly ascertained, evaluated, and tested. Again, on this principle (the 9th), the judge spelt out the necessity of getting independent evidence. He also stressed that there should be a structure to the audit to ensure that the appropriate independent audit evidence is gathered and that someone with the appropriate audit experience and competence obtains it.

The judge said: *'There are three essentials that must be met before an auditor can reasonably rely on the company's system of internal control. First, there must be a proper inquiry to ascertain the company's system. This would include ascertaining such features as indicate the strengths and weaknesses of the system and hence its reliability. Second, there must be an appraisal of it, in that a person of sufficient auditing competence should make a decision as to the extent, if any, that the auditors can sufficiently rely upon it. He should decide what procedures should be adopted to*

check that it is operating as intended and what other conditions should be met before reliance can be placed upon it. Third, there must be a testing of its operation. All these essentials may call for revision in the course of the audit. For example, because of the result of testing it might be necessary to make a decision in the course of the testing to extend the testing or even not to rely on the system'.

The underlining is to emphasise the key points that have implications for APs. This part of the judgement reinforces the earlier statement about the importance of independent evidence. It can be interpreted to mean that the auditor must properly understand the business activities (systems) being audited by the AP procedure. An AP presumes that a certain structure underlies the figure in the financial statements being audited. The auditor needs sufficient experience and skill to be able to make that judgement. To do that requires a good appreciation of the relations with the set of financial statements. The auditor must have sufficient experience, competence and knowledge to be capable of properly interpreting the outcome of the application of the AP.

4.7.2 The Court case and the Statements of auditing Standards

A summary of the seven principles that are particularly pertinent to APs is set out in Table 4.2, together with an interpretation for APs. There are a number of conclusions applicable to APs that we can take from the judgement in the case of *Pacific Acceptance v Forsyth and* Ors (1970). This section summarises those and compares them to the relevant Auditing Standards.

If there is material error in the line item of the financial statement being audited using an AP, that AP must be robust enough to have a high probability of correctly signalling such error. The converse is also true. If there is no material error, the AP should have a high probability of signalling that. The procedure cannot otherwise be interpreted as being sufficient, relevant and reliable (Points 2, 3, and 4). In general the auditor should not rely on the work of others without undertaking extensive testing to prove its worth (Points 5 and 6). In particular, management representations at the outcome stage of an AP should be carefully considered and checked against other information. APs undertaken by others, including other auditors or management, are only admissible if the work has been tested.

Judgement	Reference[6]	Interpretation for APs
The procedure must be reliable enough to find error.	Point 2	APs must have a high probability of detecting material error, should it exist.
There is a duty that the auditors' procedures should pay regard to the possibility of fraud. That should not be judged with the benefit of hindsight	Point 3	The design of the AP should be such that it has a reasonable chance of detecting material error including fraud.
Should obtain sufficient, relevant and reliable information.	Point 4	The AP must be capable of signalling material error, if present, and the opposite if that is true. It should do both reliably.
It is not a defence when something goes wrong in the audit to say that management misled me and so it is their fault.	Point 5	The auditor must not solely rely on management. Importantly in the AP process, when seeking explanations from management for variations from the auditor's expectation value the strategy of explanation and counter explanation should be developed and applied.
Reliance on sources of evidence from other parties is an aid and not a substitute for an auditor's procedures.	Point 6	The use of APs undertaken by the management of the audited entity cannot replace the auditor's own work. It is important since it re-enforces the point that auditors must control their audit procedures and evidence.
Before placing audit reliance on internal controls, these must be thoroughly ascertained, evaluated, and tested. Again on this principle, the judge spelt out the necessity for the person making the decisions resulting from the AP to have sufficient auditing competence.	Point 9	It is important in the context of APs since, like point 4 above, it emphasises the need for independent reliable evidence. It also emphasises the necessity to have someone of sufficient audit competence. The SASs recognises this. The judgement also recognises that the auditor may have to re-appraise earlier assumptions in the light of evidence that they obtain.
Carefully plan, supervise and review all work, their own and that of subordinates.	Point 10	A member of the audit team with the appropriate experience and knowledge of the methodology should administer APs.

Table 4.2: Judgements from of *Pacific Acceptance v Forsyth and* Ors (1970) 90 WN (NSW) 282 *that affect the use of APs.* [Source: Godsell (1991)].

In point 10 (Table 4.1) of his judgement the judge implied that auditors should carefully plan, supervise and review all of their work, including work performed by subordinate staff. It is particularly relevant to APs since their efficiency and effectiveness depends on the knowledge of the business and the environment. That knowledge depends on the experience and competence of the staff. So managers and audit partners who have the wider experience and exposure to a range of audited bodies

[6] Reference to Godsell's (1991) Book.

are required by SAS dealing with Quality Control, SAS 240, to ensure that *'audit work is assigned to personnel who have the degree of technical training and proficiency required in the circumstances'*. If junior staff are used, the supervision of them is paramount: *'sufficient direction, supervision and review of work at all levels is carried out in order to provide confidence that the work performed meets appropriate standards of quality'* (SAS 240: Quality Control for audit work).

An immediate question is whether the Statements of Audit Standards issued in 1995 address the other matters that were raised by the judgement. For example, the requirement for the AP to be capable of discovering error, points 2 and 3 in Table 4.2, is contained in SAS 410 on APs: *'however, reliance on the results of APs depends on the auditors' assessment of the risk that the APs may identify relationships as expected whereas, in fact, a material misstatement exists'*. SAS 110 (Fraud and Error) also makes it clear that auditors must design procedures that have a reasonable chance (high probability) of detecting error. The SAS says *'Based on their risk assessment, the auditors should design audit procedures so as to have a reasonable expectation of detecting misstatements arising from fraud or error which are material to the financial statements'*. (SAS 110.3)

Points 4 and 5 in Table 4.2 imply that the auditor cannot rely solely on management but they should do some substantive work. This is set out in SAS 300 (Accounting and Internal Control Systems), paragraph 54: *'the assessed levels of inherent and control risks cannot be sufficiently low to eliminate the need for auditors to perform any substantive procedures for material account balances and transaction classes. However, these substantive procedures may comprise only APs where such procedures provide sufficient appropriate evidence'*. SAS 440 (Management Representations) makes the point that representations by management cannot be a substitute for other audit evidence that auditors expect to be available. *'If auditors are unable to obtain sufficient appropriate audit evidence regarding a matter which has, or may have, a material effect on the financial statements and such audit evidence is expected to be available, this constitutes a limitation on the scope of the audit, even if a representation from management has been received on the matter. In these circumstances it may be necessary for them to consider the implications for their report'* (SAS 440, paragraph 15). The SAS continues with the direction: *'If a representation appears to be*

contradicted by other audit evidence, the auditors should investigate the circumstances to resolve the matter and consider whether it casts doubt on the reliability of other representations' (SAS 440.4).

The qualified reliance on other auditors is spelt out in SAS 510, which deals with the relationship between principal auditors and other auditors. For example: *'when using the work of other auditors, principal auditors should determine how that work will affect their audit'* (SAS 510.1). The 'principal auditors' are the auditors with responsibility for the audit. The reliance on experts is dealt with in SAS 520 (Using the work of an expert): *'When using the work performed by an expert, auditors should obtain sufficient appropriate audit evidence that such work is adequate for the purposes of the audit* (SAS 520.1*)*. The SAS defines an expert to be *'a person or firm possessing special skill, knowledge and experience in a particular field other than auditing'*. SAS 520 continues: *'When considering whether the expert has used source data which is appropriate in the circumstances, the auditors may consider the following procedures: (a) making enquiries regarding any procedures undertaken by the expert to establish whether the source data is sufficient, relevant and reliable; and (b) reviewing or testing the data used by the expert'*. These taken together address points 6 and 9 in table 4.2

The SASs also address the concerns at points 6 and 10, Table 4.2, about supervision. In SAS 240 (Quality Control for audit work) there is a requirement for proper supervision: '*any work delegated to assistants should be directed, supervised and reviewed in a manner that provides reasonable assurance that such work is performed competently*'. The SASs make clear that all staff working on the audit, including experts, are assistants: *''assistants' means personnel involved in an individual audit, including experts employed by the auditors other than the audit engagement partner*' (SAS 240).

Thus, the points that Justice Moffit set out in the Pacific Acceptance v Forsyth and Ors (1970) case are addressed by the Auditing Standards. The issue is the interpretation of those Standards. We will address the subject of interpretation once we have considered the surveys of auditors reported in Chapters 5 and 6.

4.8 The DTI Reports and Managerial Representation

As described in Chapter 3, an important part of the AP is to investigate the representations received from management about differences between the auditor's explanation and the account figures. The Concise Oxford Dictionary (10th Edition) defines representation to be *'statement made to an authority to communicate an opinion or register a protest'*. In the context of APs that 'authority' is the auditor, and the explanations are the management representations. Points 4 and 5 of the judge's conclusion in the *Pacific Acceptance v Forsyth and* Ors case involved management representation. The auditor must interpret and audit management's representations and observations on the findings of the auditors.

There have been other rulings in the courts. In the Kingston Cotton Mill fraud case (1896), the argument was successful that auditors are entitled to rely on management's representations. Auditors frequently cite this case as a justification for relying on management's representation. In contrast to that case, the Thomas Gerrard (1967) judgement found that auditors were liable for not detecting a major management fraud based on the inclusion of fictitious stocks and the manipulation of end of year procedures. In this case the auditor did query the end of year figure, but accepted management's representations. The implication is that the auditor must be diligent and use their knowledge to question and challenge the audited body, particularly if they are going to take as evidence management's representations. Thus interpretation of managerial representation plays a key part in the AP process.

We now look at whether reports into company failures by the Department of Trade and Industry (DTI) provide a source of evidence on where auditors have made the wrong, or inappropriate, conclusions about the financial statement of an organisation. Using the legislation contained in the Companies Acts of 1947 and 1948, the DTI may appoint Inspectors to investigate the affairs of a company. Under the provisions of the 1985 Companies Act the Inspectors may be appointed for different purposes: to investigate the affairs of a company, to investigate ownership, or to investigate share dealings by the Directors or their families. The reports are usually very detailed and appear many years after being commissioned. Although the Inspectors are not appointed to investigate auditors, many investigations uncover cases of malpractice and irregularities involving the auditors of companies. These often have financial

implications. It is important to remember that these reports only deal with failures that the Inspectors are directed to investigate, and not all failures.

Boys (1997) reviewed the DTI Inspectors' reports covering the period 1971-1995. He provides some interesting evidence on where auditors' work failed to detect problems within the financial statements of organisations, and on the impact of such reports on auditing practice. Of the eighty-eight reports published in the period 1971 to 1995, forty-five (51%) had significant comment on the work of auditors.

The issues relating to auditors can be illustrated by a simple analysis of Boys' findings. Of the forty-five reports with reference to auditors, thirty-three contained references to problems with the evidence obtained by auditors. Boys cited seventy-four references in those thirty-three reports.

There were recurrent themes of auditors being deceived by management, of substandard judgement by auditors, of poor documentation and poor application of audit processes. Auditors were criticised for allowing themselves to be deceived through misrepresentation by the Directors and senior management. Other judgements of auditors were also criticised. Instances included accepting management representation without inquiry, not obtaining independent evidence of valuation, relying on inadequate evidence or inappropriate tests, carrying out insufficient audit work and failing to document their work. Table 4.3 below summarises the nature of the findings (note that some reports had more than one category of criticism). The main implications are that it was not the particular techniques that were the cause of the auditor's failing, but rather the auditor's application of the procedures, including inappropriate acceptance of management representations and evidential issues.

Comment on the Work of Auditors	**Number of reports**	**Proportion of the 45 Reports**
Number of reports with significant comment on the work of auditors	45 Reports	100%
Misrepresentation by Directors or Management	17 Reports	38%
Problems with evidence	33 Reports	73%
Use of a particular audit technique	2 Reports	4%

Table 4.3: A summary of the DTI Inspector's comments on the work of auditors. [Source: Derived from P Boys (1997)].

Although there were many instances of misrepresentation by management to auditors in Boys, there were only two comments about the use of particular audit procedures. One related to sampling: *'there was no scientific basis for the selected sampling'* (The 1978 DTI Report on Roadships). This particular sample related to creditor circulation. The other was a criticism about the poor application of an AP. The AP involved the calculation and investigation of gross profit margins to test whether stock had been overvalued at the end of the accounting period. The Inspectors found that there was clear evidence that the AP was improperly applied. The accounts showed a gross profit margin of 16.5% on all products sold. The auditor's AP predicted a gross margin of 10%. They failed to get adequate explanations for the difference of 6.5%. Since the auditor relied upon an AP in the testing of this balance and collected no other substantive evidence (e.g. by sampling) the failure of the auditor to properly carry out the AP meant that inadequate evidence was obtained, and the Inspectors were critical of that. This was a failure of the auditor to develop a proper expectation that took account of all the relevant variables and to have a strategy for obtaining reliable explanations.

The evidence we have from Boys (1997) suggests that, in the past, auditors have been too willing to accept management representations. For APs, the implications are serious as explanations are sought from management to explain any difference between the expectation-value and the figure being audited. This is a key part of the final part of the AP process: the outcome rule.

For the example above, from the 1978 DTI Report on Roadships, there is an alternative explanation to the one of management representation. It is one that, unfortunately, we cannot test directly. Suppose that the outcome-value set as the criterion for assessing the outcome of the AP was greater than 6.5% and so the auditor failed to seek adequate explanations because they considered the difference of 6.5% to be immaterial. For example the difference between the auditors prediction (10%) and the audited value was 6.5% (16.5% - 10%). If the outcome-value applied to the test was 8%, then the observed difference would have been judged 'not significant' and the figure being audited would be accepted. This is one potential explanation of why APs are perceived by auditors to be effective – they give positive signals. Unfortunately that might be, as in this example, even when there are large differences. That can arise from a poor

investigation rule. We shall return to this and examine in Chapter 8 whether many of the investigation rules adopted by auditors give false-positive or false-negative signals.

4.9 The investigation and outcome rules

In section 3.8 we described an AP as comprising two rules: the investigation rule and the outcome rule. A key part of the investigation rule was building the model to predict an expectation-value. The development of an expectation-value for the AP is a key stage in the whole AP process. According to Lev (1980): *'the crucial stage of the analytical review process is the generation of expected, or reasonable, values of financial statement items'*. Developing an expectation for a value in the financial statements requires the auditor to model the business processes that produce the value. Echoing the court ruling and the SASs, an essential part of that is understanding those business processes: '*Analytical review procedures rely upon an auditor's ability to generate an expectation of an account based upon knowledge of a company's business and operating environment*' Knechel (1988a).

The necessity to consider the appropriateness of the expectation value was spelt out by Albrecht and McKeown (1976) who described APs as seeking to answer the question. '*Are the current financial statement balances reasonable when considered in the light of what one would expect them to be, based on previous years' (months') balances? If not, why not*?' Implied in such a definition is a concept of measurement of the acceptability of the auditor's estimate; this is absent in the auditing standards of that time. There is a development of the expectation, the comparison and the closeness (precision) of the expectation to the audited balance. Kinney (1978), in one of the early pieces of research on the effectiveness of APs, provides a description of the whole process. '*In any analytical review, the auditor compares the client's reported balance (or ratio) with the auditor's assessment of the likely true (audited) balance. This assessment or conditional expectation may incorporate the audited balances of the client for prior periods and/or structural data such as current and past data for the economy and industry, current unaudited data from related accounts and independent internal records such as production statistics. If the client reported values are 'close' to the auditor's conditional expectation, the auditor's confidence in the validity of the reported balance is increased. That is the auditor assesses a lower probability of the existence of material accounting error'*.

Until recently the explicit need to develop expectations and consider the precision of those expectations has not figured in the auditing standards in the UK. Blocher and Patterson (1996), who drafted much of the current guidance on APs in the USA, set out a paradigm for an AP and the key considerations affecting the precision of an expectation. The paradigm involves three steps. They are the development of an expectation, comparison of that with the recorded balance and a decision against some significant value, with some consideration of the precision of the expectation value. These are the equivalent of our investigation and outcome rules. The key considerations affecting the precision of an expectation are: the methods, the reliability of the data, the aggregation of the data and last, the predictability of the relationship. Of these, the first and last are equivalent to the quality of the model, while the second and third reflect data quality.

Another consideration is the use of confirmatory information. These recognise that effectiveness and reliability of such work will depend on the diligence, skill and knowledge of the auditor. The AP is one of many signals and the interpretation of that signal, in the light of the other evidence, is one of the important skills highlighted by the judge in the Pacific Acceptance v Forsyth and Ors case. When talking about the care required from auditors in placing reliance upon internal controls, the judge said: '*a person of sufficient auditing competence should make a decision as to the extent, if any, that the auditors can sufficiently rely upon it*'. The 'it' in this quotation is the internal controls, although the same logic applies to interpreting the results of an AP or all the different sources of evidence about financial statements.

For an AP, the final stage, the decision about the success of the AP, requires an assessment as to whether the expectation value, including the associated precision, are sufficiently close to the value in the audited statements. This was described by Kinney (1978), as the closeness of the expectation value to the reported balance. This process will be considered in more detail in Chapter 7.

4.10 The Explanation Effect

Once the auditor has calculated the expectation-value and has compared it to the book-value using the outcome-value as a criterion, the next decision is whether the difference is significant. The associated question is what is the explanation for it. At this point the

auditor is in a dialogue with the management as part of the process of establishing whether there is a plausible explanation for the observed difference. For example, the management might bring to the auditor something of which the auditor was unaware.

As was described above, Boys' (1997) report on the work of the DTI inspectors points to another potential problem for auditors: a willingness to accept too readily management representations about information concerning the financial statements. This may take the form about explanations for the position on the stock of an organisation, as in the Thomas Gerrard (1967) judgement. The work of the Joint Monitoring Unit of ICAEW (JMU) revealed continuing problems with the unwillingness to challenge of management representations. *'It is not enough merely to accept a written representation from the directors'* (Garvey and Dietz (1997)).

This willingness has implications for APs in the obtaining proper explanations for the difference: $|\hat{Y}_T - Y_B|$. Anderson and Sechler (1986) highlighted the problem in another group of workers and called it the explanation effect. This is the willingness of all humans to accept plausible explanations to a question, without questioning the accuracy of those statements. Although one might argue that auditors are trained to overcome this trait, Libby (1985) found in experiments with audit staff from two major audit firms that they did in fact accept such representations. Certainly, the criticism illustrated in Boys (1997); that auditors allowed themselves to be deceived through misrepresentation by the Directors and senior management, implies that auditors like other humans have a problem with the explanation effect.

An illustration of the theory developed to mitigate the problem is contained in the work of Anderson and Sechler (1986). They carried out experiments to ascertain whether the explanation effect could be minimised by getting the decision-maker to postulate both explanations and counter-explanations. If the decision-maker is the auditor, one of their findings was that the auditor needed to understand fully the relationships between the variables that underlie the expectation value. They then need to consider all the alternative explanations that could explain a difference between that expectation and the value in the financial statements. *'The main error leading to the explanation effect is not in using the availability of plausible causal explanations in judging the probable relation between two variables. Rather, the (explanation) error seems to be grounded in people's inability (or unwillingness) to see that the availability of a particular*

explanation may have been due to factors unrelated to the truth of the explanation, and that equally plausible causal explanations could be generated for alternative or opposite relations'. They concluded that '*If people typically considered all possible alternatives before making important decisions, the explanation bias might be relatively unimportant; the various counter-explanations would tend to leave the decision maker relatively unbiased'*. Koonce (1992) has extended this work. She carried out experiments using auditors looking at avoiding error when obtaining explanations for the differences between expectation values and the account value. She found that '*the natural order for comprehension may be to explain and then (possibly) counter-explain*'.

To minimise the explanation effect, we propose a sequence of 'rules'. They are based on the theory of explanation and counter-explanation and are derived from the work of Anderson and Sechler (1986). They are:

1. Identify potential alternative explanations before doing the AP;
2. Carry out the AP;
3. Review the alternative explanations before talking to the audited organisation;
4. Seek explanations from the audited organisation; and
5. Substantiate those explanations.

As a result of research for this thesis, the author proposed that those rules should be implemented within the NAO. That was possible because the author was the Director responsible for statistics and research. Following agreement from the senior management, in September 1999, these rules were implemented through a new chapter on APs in the NAO Audit Manual and training courses. In the audits following the introduction there was an improvement in audit quality. A quality assurance process adopted by the NAO monitored that improvement. Each year a number of completed audits are subjected to peer review at the level of Director (the equivalent of partner in the private sector audit firms). The purpose is to ensure that the audits comply with the Auditing Standards and NAO audit policy.

However, the introduction of the rules was not set up as an experiment, and they were not the only change introduced by the audit chapter. Thus it is not possible to isolate the impact of rules within the new strategy for the application of APs. However, discussion with Directors in charge of the audits confirmed that the rules focused the audit staff on the issues surrounding the elicitation of explanations for any difference ($\hat{Y}_T - Y_B$).

4.11 Summary and Conclusions

This chapter has illustrated that the auditor is required to design the audit procedures to be such that, if material error exists, then the procedures will have a reasonable expectation that it will be revealed. The Courts and the Auditing Practices Board have set out the principles for the conduct of an audit. The latter are promulgated in the Statements of Auditing Standards. For example, the Auditing Standards require that some level of substantive testing should be done. The Standard on accounting and internal control systems and audit risk assessments (SAS 300) directs that: '*The assessed levels of inherent and control risks cannot be sufficiently low to eliminate the need for auditors to perform any substantive procedures for material account balances and transaction classes. However, these substantive procedures may comprise only APs where such procedures provide sufficient appropriate evidence*'

Unfortunately there is evidence that the expected high levels of competence are not always shown in the implementation of the audits. The DTI Reports highlight that auditors sometimes use inappropriate methods or rely on management representation without sufficient scepticism (Boys 1997). There was evidence from Humphrey and Moizer (1990)) and Garvey and Dietz (1997) of auditors' failing to understand the procedures they implement.

All of these issues have serious implications for the use of APs. APs are a collection of methods requiring judgement and skill. If inappropriate methods are selected, the question is whether such methods will have low probabilities of false-negative and false-positive signals? Alternatively, is the implementation of APs carried to sufficient standards? For example, are the outcome-values so high that significant error would not be detected? The importance of these questions depends on the extent of the use of APs and the type of APs used. For instance, if APs are little used then the question, although interesting, will have no immediate practical impact. On the other hand, if APs are used

a great deal, the issues of the type of AP and the way they work are critical to the quality of the auditor's work. Also of interest will be auditors' motives for using APs and their perceptions of them.

The next stage of the thesis considers the extent of the use of APs and the types used. To collect that information, the next chapter will discuss a number of published surveys about the use of APs. That is followed in Chapter 6 by reporting a new survey that has attempted to fill some of the gaps in the knowledge about the use of APs. Subsequently we report experiments into the effectiveness of APs. We then examine such APs to see if they inherently have low probabilities of false-negative and false-positive signals.

Chapter 5: The contribution of surveys to our knowledge of APs

5.1 Introduction

Since 1984, a number of published surveys have recorded the auditor's use of, and perceptions about, APs. The subjects covered included what types of APs were used, when in the audit cycle they were used, why they were used and the extent of their use.

Because of the large number of auditors in the UK and the USA, surveys are necessary if their views are to be accurately assessed. The majority of the reported surveys have used postal questionnaires, although there have also been some face-to-face interviews. In considering the surveys it is necessary to bear in mind the two categories of error that arise from surveys: sampling and non-sampling error. The former is a direct consequence of looking at a sample rather than at the complete population: it is the error associated with inferring the population value from a sample. An extensive theory exists for estimating such errors.

Non-sampling error is harder to control. The sources of such error include: an inadequate record of the targeted population (sampling frame), inappropriate sampling methodology, poorly designed questionnaires, inappropriate use of sampling methodology (e.g. quota sampling), interviewer bias, recording and measurement errors and non-response problems. If they are not addressed, then inferences about the population could be, and probably will be, misleading. As we will see, all have been issues to some extent in one or other of the surveys.

The rest of this chapter looks at a number of surveys carried out in the USA and two in the UK, addressing the issue of non-response where relevant. The first four surveys were published in the USA over the period from 1983 to 1994 and the fifth and sixth in the UK in 1997 and 1999. As non-response was an issue for these surveys, the review will look at the data collection methodology as well as providing a précis of the key findings.

5.2 The work of Biggs and Wild (1984)

The stated aim of the Biggs and Wild (1984) survey was to obtain an insight into practising auditors' experience in using APs. The population targeted by the survey was practising auditors from the large firms in the USA. At that time, there were eight large

firms, four of which co-operated in the study. The survey was confined to one large city in the USA. The sampling was organised by a partner within each of the firms. There were 127 respondents (a claimed response rate of about 80%).

A short questionnaire was used, based around three groups of questions. The first group of questions sought information about the percentage use of scanning, ratios, regression and time-series models. The second group asked about the percentage of errors *initially detected* by an AP and the third group of questions explored the number of years of experience of each respondent. Biggs and Wild used this latter piece of information to demonstrate that respondents to their survey had the same profile of experience as staff at the national firms. They then used this to justify their results as being representative of auditors in the large firms across the USA.

Biggs and Wild reported that judgmental procedures such as scanning and ratio analysis were the most commonly used techniques. The percentages of use are set out in Table 5.1.

	Scanning the Data	**Ratio Analysis**	**Regression Models**	**Time Series Models**	**Other**
Percentage Use	96%	89%	11%	8%	28%

Table 5.1: Percentage use of APs by type of Technique. As can be seen, judgmental procedures are used most frequently and quantitative procedures infrequently.

Another finding of the survey was the auditors' perception that the judgmental APs were by far the most valuable. This view was consistent with the use of the different techniques: the simple qualitative APs were the most popular. Biggs and Wild also concluded that the perceived ability of a technique to detect errors was distributed evenly amongst the techniques, their tests showed that none were significantly better than the rest.

One particular finding was that APs initially signalled 42% of all errors detected by auditors. Table 5.2 reports the statistics relating to this point estimate; they imply that there was a wide divergence of views. This finding is significant for the history of APs in that it has echoes in all the subsequent surveys.

Median	45%
Mean	42%
Standard Deviation	26%

Table 5.2: The percentage of errors that auditors claim are initially signalled by APs. The results indicate that there is a wide divergence of views.

The survey was for economic reasons, restricted to one city and the audit staff of four of the large firms. What is also significant is the distribution within each firm of the questionnaires by a partner. One can understand the logistics of this, but was there an incentive to distribute the survey documents to the better staff? Biggs and Wild tested the distribution of staff experience and found the sample to be consistent with that of the population of firms but, although this may give some degree of comfort, that may not be the important criterion. Another potential problem is that the sample was small. A three-stage sample was used where the first sample from 4 firms from 8 (not at random), the second sample was 1 city and the third sample was 127 staff. Thus the variances of the estimates will be large (larger than stated in their results). The implication is that the point estimates might not reflect the true position since they were not weighted to reflect the sample plan. This concern is of importance because the survey is one of the first to put into the audit folklore the statistic: 40% of all identified errors are initially signalled by APs.

Biggs and Wild's method of selecting the sample must caution against an extrapolation across all auditors, or even across all auditors in the USA. Despite that, this survey is important in the history of APs. They found that simple, qualitative techniques dominate the AP scene and that a large portion (40%) of error found during an audit is believed to be initially signalled by APs.

5.3 The Daroca and Holder Survey (1985).

This survey is different to the others described in this chapter. Its target population was audit firms, whereas the other surveys targeted individual practitioners. Daroca and Holder (1985) tried to avoid bias by careful design and consultation with several practitioners during the design process. The questionnaire was distributed to 1600 USA audit firms by the American Institute of Certified Public Accountants, who controlled the timing and mailing of the questionnaires. Despite their best efforts the non-response rate was 83%. Daroca and Holder established by looking at envelopes that were returned as undelivered, that the mailing of the survey had been done over several weeks. They did not control the mailing schedule and did not know what it was. They give this as the reason for not being able to test for non-response bias.

The 17% response rate consisting of 269 replies has not attracted much attention; the question of whether the resulting sample is representative of the situation across the USA is not addressed. It may be that only firms using APs replied and thus the results significantly overstate the usage of APs in the population. Given that this work was published in 1985, it is now difficult to determine whether other checks could have been done to see if there was any bias. It would have been useful to look at the profile of firms responding against those who did not, and to compare those with the population profile. Another criterion might have been with respect to the size of firms.

Daroca and Holder reported that there was extensive use of simple APs, using the data contained within the audited financial statements. Such procedures typically involved simple comparisons or 'standard ratios', and auditors reported that they were applicable in over 60% of all audits (the responses ranged between 60% and 90%). More complex procedures such as regression and other statistical techniques were not considered to be applicable to many audits: only 16% of firms responding claimed to use them. This survey provides an early example of a worrying feature of the application of APs: that many seem to be used by auditors without inquiry as to whether they are appropriate for the particular audit. Daroca and Holder hypothesised that certain APs were used merely because they had been utilised in the past without considering the diagnostic ability of those procedures. They considered that: *'some highly focused procedures which would seem to spotlight specific error identification are used only infrequently'*. In other words, auditors do not focus on the appropriateness, predictability or power of the procedure they apply.

5.4 The work of Tabor and Willis (1985)

This was an in-depth study of audit managers working for one of the big audit firms. For logistical reasons, the sampling was constrained to one firm that agreed to provide the researchers with the staff time to participate in the survey. Seven managers were chosen for their experience: one had been with the firm for five years; the others had been with the firm for between seven and nine years. This was not intended to be a random sample for the purpose of extrapolation. It provided an in-depth view of how one small group of auditors used APs within the context of the audit strategy of that firm. It also provided information on auditors' perceptions of APs. Each manager

discussed two audits that they had recently completed. The interviews were based around their audit working papers.

Tabor and Willis (1985) found that APs were extensively used during planning (on average over 30% of the audit hours devoted to planning) and at the substantive stage of an audit (on average over 33% of the audit hours devoted to substantive procedures). In contrast, APs were used infrequently at the review stage. An interesting insight from this face-to-face survey was the feedback on what the seven managers thought would be significant developments in the years after 1985. Three thought that the use of computers would make APs easier to use, and five thought that statistical methods would play a bigger role. Increasing fee pressure was also considered to be an important factor. Thus in 1985, a picture emerges of APs being an important tool, but not then exploited to their full potential.

This survey provides an insight into the work of a small group of auditors. From the survey it is possible to gauge some measure of the increase in the usage of APs and of the procedures used within the audit firm for which the managers worked. Tabor and Willis reported that over the period covered by the survey (1978 to 1983), the use of APs at the planning stage had not changed much, and that simple procedures were used. Over the period there was a reported 11% increase in the use of APs at the substantive stage of the audits. A mixture of simple methods through to regression-based methods was used, but the simple methods predominated. This is consistent with the findings of Albrecht and McKeown (1976) and Daroca and Holder (1985).

5.5 The work of Ameen and Strawser (1994)

This survey was carried out in the early 1990s and reported in 1994. The intention was to select a sample of practising auditors. For economic reasons, Ameen and Strawser (1994) restricted the survey to two cities in Texas, USA. They claimed that the information they had was about qualified accountants listed by the relevant professional accountancy body in two large cities in the southwestern states of USA. Actually upon enquiry it turns out that both cities were in Texas. Strictly speaking, the inferences from their work are confined to that state, although we know that the big firms operate on a national and a global scale. Thus, although there are local interpretations of the strategy, the results relating to the firms from the big six can be extended to the USA.

Some 895 accountants were identified of whom an unknown number worked as auditors. Ameen and Strawser did not know which of these accountants were practising as auditors. They therefore adopted a sampling strategy of self-selection and post stratification. They sent a questionnaire to all 895 accountants, of whom 414 returned the questionnaire indicating that they were not involved in audit. As a result of two mailings, Ameen and Strawser received 190 responses from accountants who were practising auditors. Assuming that the 291 accountants (895-414-190) who did not reply were auditors, Ameen and Strawser (1994) claimed a response rate of 40%. They post-stratified their sample into the Big Six firms (100) and smaller firms (90). Since the split of the non-response information between the strata is not known, a comparison of the big six and the smaller firms cannot readily be made. Despite this, the survey provides very useful additional information about the use of APs.

Ameen and Strawser tested for non-response bias by making comparisons of the responses and other demographic information provided by the respondents to the two mailings. These revealed no statistically significant differences, and they concluded, *'Any non-response bias on the results should be minimal'*. Ameen and Strawser did not report any other analysis.

The research instrument was a questionnaire covering nine topics. The purpose was to establish the proportion of time spent doing APs, and the types of procedure used. This information was collected for the planning, substantive and review stages of an audit. The APs used were the commonly defined ones of comparison, ratio analysis, judgmental trend analysis, statistically based time series analysis and regression. There were specific questions to obtain information on the changes in the use of APs over the last 5 years and the reasons for those changes. Other than for the questions seeking to establish the proportion of usage, a scoring system was used with a range of seven points for each of the questions: with seven implying that the procedure was used extensively, and a score of one indicating that such procedures were never used.

Ameen and Strawser (1994) found that auditors still rely on simple methods. Figure 5.3 shows the relative importance of different APs ranging from very simple methods to regression analysis. There is little apparent difference between the large and smaller firms in the relative use of each of the individual procedures, but as we do not know the split between non-response of the small and large firms this may not be correct.

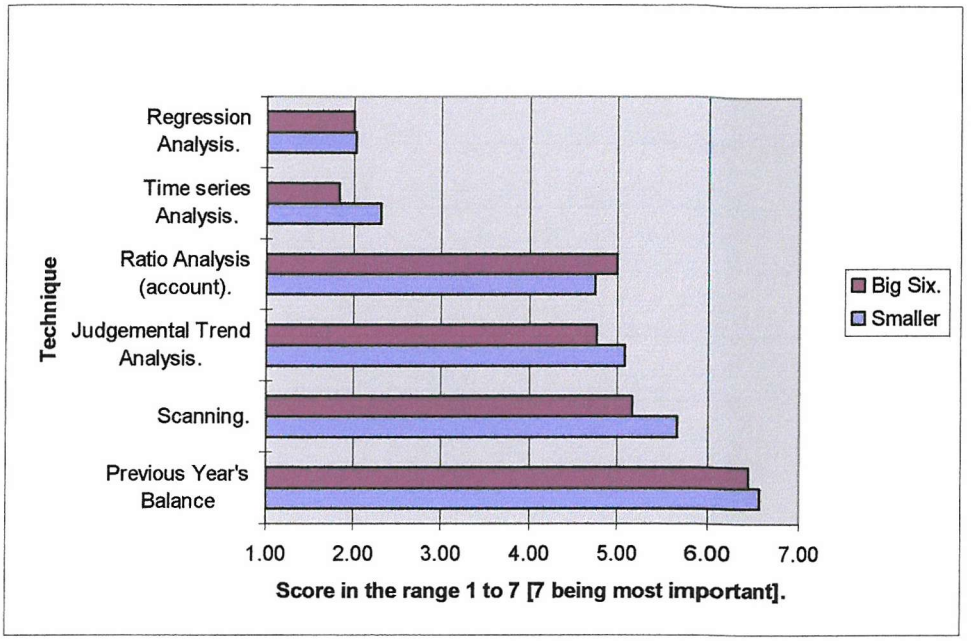


Figure 5.3: This illustrates the use of different types of APs. The chart shows that simple procedures dominate APs. A scoring system was used to compile this data, a score of seven indicating that the procedure was used extensively and a score of one that it was never used.

Differences between the audit firms are revealed when one considers the proportion of time that APs are used, relative to the overall hours spent on an audit. This is summarised below in Table 5.4.

The Time Spent On APs at Each Stage of an Audit.		
	Big Six Firms.	**Smaller firms.**
Planning Stage.	48% (32%)	37% (27%)
Substantive Testing Stage.	43% (19%)	33% (20%)
Final Review Stage.	55% (31%)	33% (27%)

Table 5.4: This shows the proportion of time spent using APs at each stage of an audit. At the 1% level of significance, there is evidence to reject the null hypothesis that there is no difference in the use of APs between the big and smaller firms at the substantive and review stages of the audit. The figures in brackets are the standard deviations. (Source Ameen and Strawser (1994)

What comes out of this is the importance of APs in the whole audit process. However, the standard deviations associated with these averages are large, reflecting the variations between firms. These are shown in the table in parentheses. The times devoted to APs by the big six and small firms to APs at each stage of the audit were compared and null hypothesis set that there was no difference in the proportions of time devoted to APs by the big six and small firms. A t-test was then carried out on the difference between the proportions of times. It revealed that there is evidence at the 1% level of significance to reject that null hypothesis for the data concerning the substantive and review stages of the audit. There is also evidence to reject the null

hypothesis for the planning stage of the audit. Thus there is a difference between the times that is not only visible, but is statistically significant.

The Ameen and Strawser survey also asked auditors for their views on what were the most important reasons for the changes over time in the use of APs with a score of 7 to indicate a large effect, and a score of 1 to indicate no effect. These are shown in Figure 5.5 below.

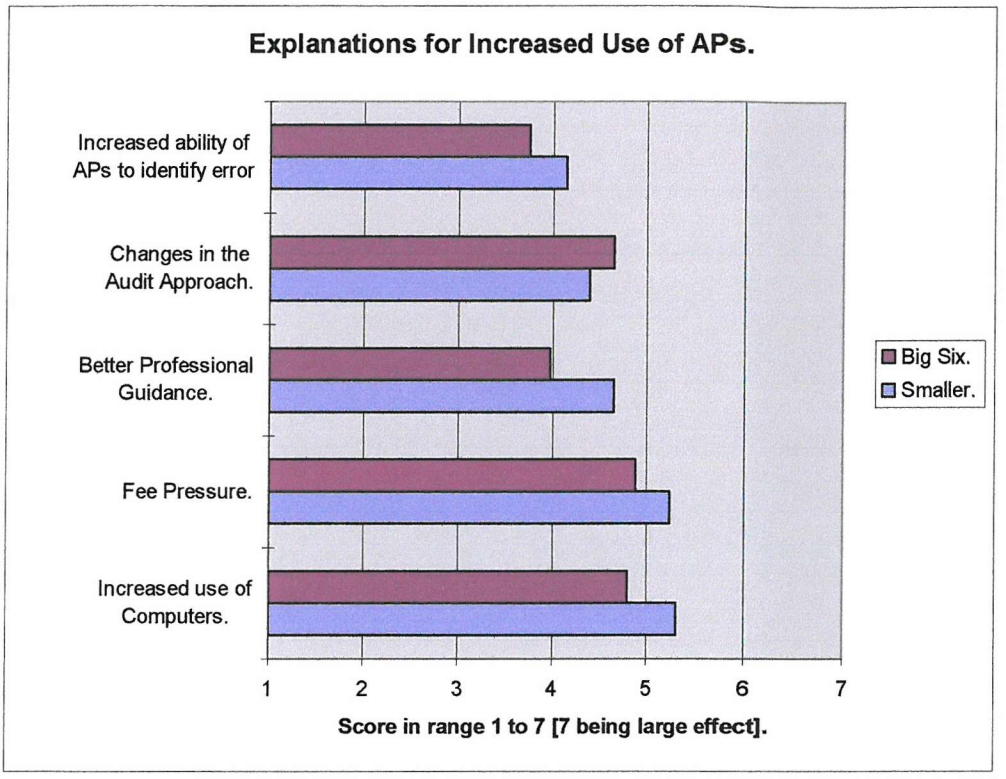


Figure 5.5: This shows the reasons given by the auditors surveyed for the changes in the use of APs. A score of 7 indicates large effect while a score of 1 indicates no effect.

Clearly all the factors were considered important in the changes in the use of APs. The increasing use of computers and fee pressure were thought to be the most important factors. This matches the prediction reported by Tabor and Willis (1985) that these would be important factors in the increasing use of APs.

Ameen and Strawser suggested that: '*the use of Analytical Procedures increases when (1) the entity's internal control procedures are considered to be effective, (2) the account has a lower risk of misstatement, and (3) the audit represents a continuing engagement.*' This conclusion supplements the earlier proposition of Wright and Ashton (1989) which was that: '*when controls are strong, procedures involving*

internal accounting data are more diagnostic: while with weak controls evidence external to the accounting records signals relatively more errors'. The logic behind this is that where organisations have good control procedures, including strong management, then the data produced by these systems will be good and the relationship between systems and their data will be stable. In such an environment the presumption is that APs will reflect the good management and signal reliably areas where things may, potentially, be wrong. On the other hand, if the systems are poor, internal data will have many conflicting signals (noise) and only APs using data external to the organisation could be effective in establishing the absence of material error.

5.6 The Fraser, Hatherley and Lin Survey.

The survey by Fraser, Hatherley and Lin (1997) was the first published large-scale survey into the use of APs in the UK. The survey targeted audit partners in overall charge of individual audits. The earlier survey by Humphrey and Moizer (1990), which interviewed experienced audit managers, looked at the audit methodology in general, whereas Fraser, Hatherley and Lin (1997) specifically investigated the use of APs. .

Fraser, Hatherley and Lin sampled 700 partners in 155 firms across Britain, and they received 366 replies from partners in 146 firms. The sample was post-stratified by size of firm: the big six, large (more than 20 partners), medium (between 6 and 20 partners) and small firms (fewer than six partners). At the time of the survey there were 6 big audit firms, since then two of those firms have merged.

5.6.1 A summary of the results of Fraser et al

The aim of Fraser et al (1997) was to obtain information on five main issues:

I. The extent of use of APs;
II. The relative use of different techniques;
III. The role of APs in evidence collection process;
IV. The perceived effectiveness of APs, and of specific AP techniques in the detection of error; and
V. The factors that have influenced recent changes in use of APs.

5.6.1.1 Extent of use of APs

Fraser, Hatherly and Lin's (1997) survey reveals that all the firms in the UK use APs extensively. At the planning and testing stages there is an association between use and firm size, with the larger firms making more use of APs. There is a marked difference between the big 6 firms and the rest. At the testing stage, the big firms use APs in over 75% of audits, whilst the large and medium sized firms use them in about 60% of audits. By contrast, the small firms claim to use them in over 40% of their audits, still a high usage. These results are shown in Figure 5.6, which clearly demonstrates the high use of APs.

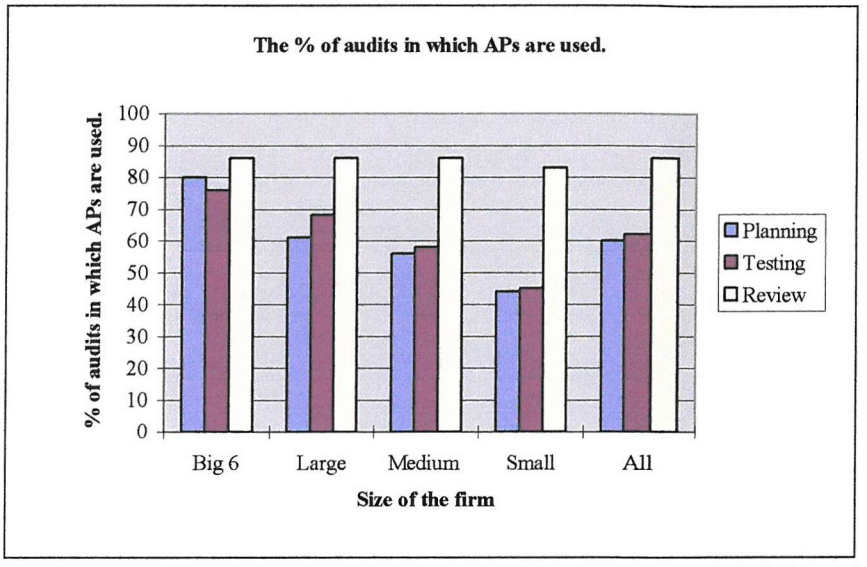


Figure 5.6: This shows the percentage of audits in which APs are used. Clearly they are used at all stages of audit. There is an association between firm size and usage: the larger the firm the greater the use of APs. [Source: Fraser et al]

Now that we know that APs are frequently used, what is the distribution of usage within the categories of audit firm (big, large, medium and small)? The survey revealed a wide range of the extent of use of APs on audits within firm category (size). Table 5.7 below shows the distribution of the use of APs by auditors in the firms: it is stratified by firm size. This is of interest because it implies that APs are not used consistently within a firm. That could reflect the different types of client, or different attitudes within the profession towards APs. Turley and Cooper (1991) identified this latter point as an issue in their survey of auditors in behalf of the ICAEW. They found that there were wide differences in the way APs were defined and specified. They also found that the amount of assurance taken from APs varied widely between audit firms. Turley and

Cooper's interviews revealed fairly widespread divergence between the theoretical definition and the practical understanding of APs. That survey is not described here because it was a survey of general audit methodology and only touched on briefly APs.

Percentage of audits in which APs used	Percentage of respondents using APs by Firm Size				
	Big	Large	Medium	Small	Wt Average
0% to 20%	4%	4%	17%	18%	10%
21% to 40%	3%	10%	14%	26%	12%
41% to 60%	18%	20%	18%	34%	20%
61% to 80%	11%	22%	16%	11%	18%
81% to 100%	64%	44%	35%	11%	40%
Total	100%	100%	100%	100%	100%

Table 5.7: This represents a distribution of usage of APs by the different sized firms. Each cell represents the percentage of respondents using AP on a proportion of the audits. For example the top left cell shows that 4% of auditors in the Big Six firms use APs at the testing stage on between 0% and 20% of the their audits. [Source: Calculated from a Figure in Fraser (1997)]

Given that APs are used extensively, a question that follows is which categories of procedure are used and at what stage of an audit. This information is summarised in Figure 5.8.

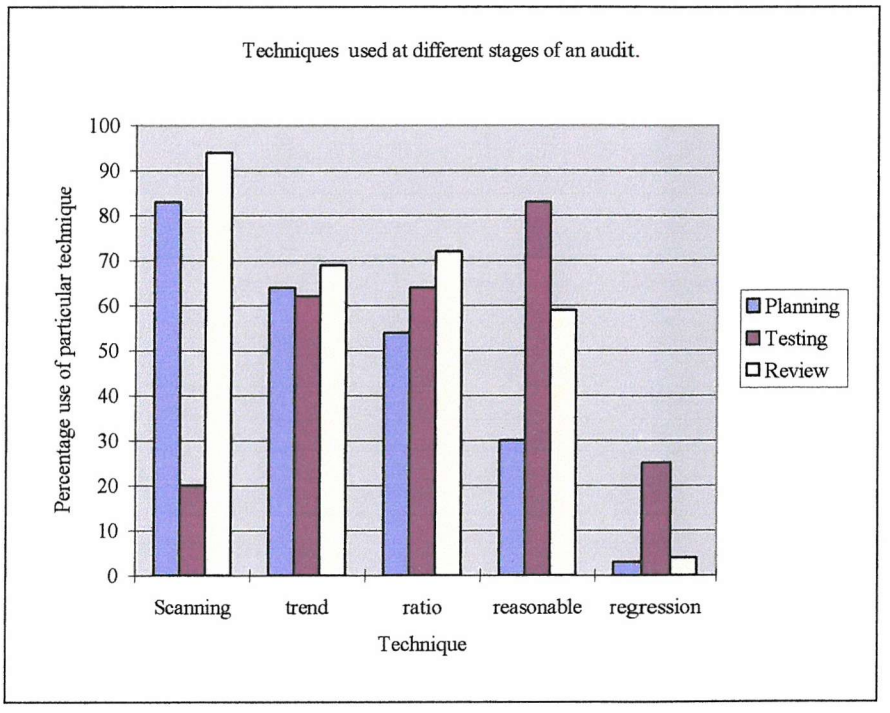


Figure 5.8: This graph shows the use of techniques used at each audit stage. Clearly, the simpler procedures (scanning, trend and ratio analysis) are used more frequently at all stages. Somewhat surprisingly modelling (e.g. reasonableness tests) is frequently used at the review stage. [Source: Fraser et al (1997)]

The survey found that the simpler procedures (such as scanning, trend and ratio analysis) predominate at all stages. Methods such as regression analysis are used only at the testing stage of audits. Somewhat surprisingly modelling procedures such as reasonableness tests are used frequently at the review stage.

Reasonableness tests as defined in chapter three involve explicit development of expectation-values and consequently are time consuming. A likely explanation is that auditors interpreted reasonableness tests to be just what the word implies: a test of the reasonableness of the final financial statements. That would be consistent with the Auditing Standards: *When completing the audit, auditors should apply analytical procedures in forming an overall conclusion as to whether the financial statements as a whole are consistent with their knowledge of the entity's business. (SAS 410.3)*

5.6.1.2 Relative use of APs

The next logical question is whether there is a difference between the firms of different size in their use of APs at the three stages of an audit. This information is shown in Figures 5.9 to 5.11. As can been seen from all three Figures, the clear message is that simple procedures such as scanning, trend analysis and ratio analysis are used extensively at all stages, by all firms. The techniques that require the development of explicit expectations such as reasonableness and statistical modelling, are used more frequently at the testing stage than the planning stage, as is demonstrated by comparing Figures 5.9 and 5.10. At the substantive stage (Figure 5.10) there is some association between the method use for an AP and the size of the firm: generally the larger the firm the more use of all types of AP. The results for the review phase of the audit are shown in Figure 5.11.

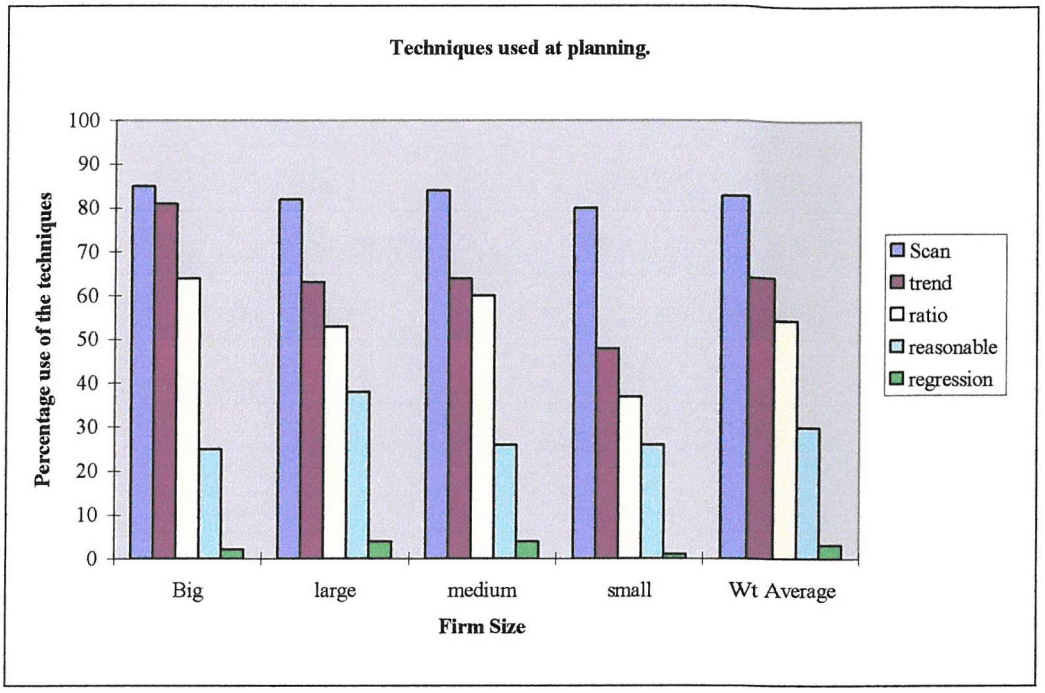


Figure 5.9: This graph shows the techniques used at the planning stage of an audit for the different sized firms. Scanning is the most common technique to be used followed by trend, ratio and reasonableness. Regression is rarely used. [Source: Fraser et al]

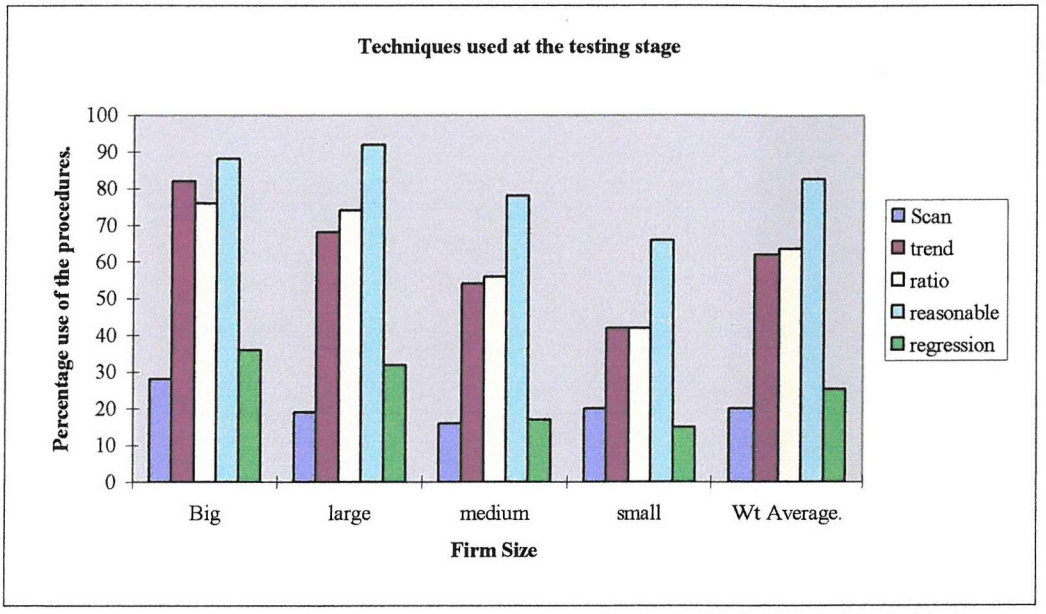


Figure 5.10: This graph shows the reported use of different techniques at the testing stage of an audit by the different sized firms. Clearly, trend and ratio analysis and reasonableness tests are used most frequently. Scanning analysis is widely used. Regression analysis is used as frequently as scanning. [Source: Fraser et al]

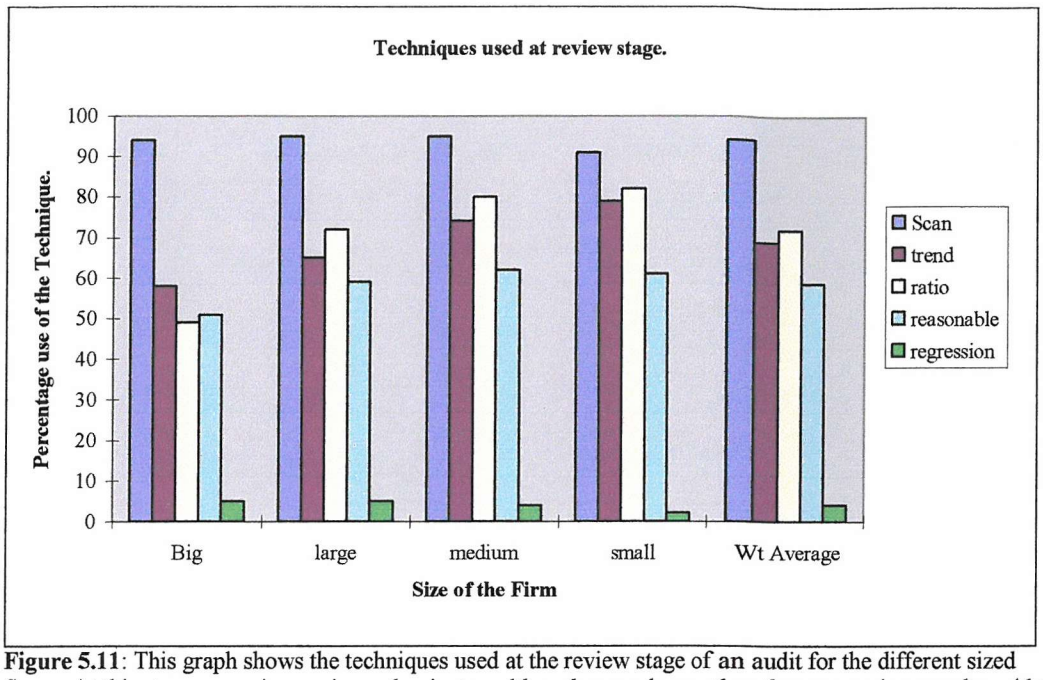


Figure 5.11: This graph shows the techniques used at the review stage of an audit for the different sized firms. At this stage, scanning again predominates, although procedures other than regression are also widely used. [Source: Fraser et al (1997)].

Looking at the results for the review phase of an audit (Figure 5.11) we see that regression is not used much. As we remarked earlier, the extensive use of reasonableness tests is surprising, and probably reflects a different definition of reasonableness methods by the auditors. The extensive use of the simple methods is consistent with the planning phase of the audit.

5.6.1.3 Perceived effectiveness of APs and their role in reducing other tests

Fraser et al asked the auditors to rank the effectiveness of the different procedures. These findings are at Figure 5.12, from which it is clear that auditors regard the simpler techniques as the most effective. If these are compared to the rankings of academics, which were gathered by Kinney (1978 and 1979), the perceptions are different: particularly with regard to scanning and regression. Academics view regression to be the most effective, while auditors perceive it as the least effective, with scanning being viewed as the least effective by academics and the most effective by auditors. One of the likely explanations for this apparent contradiction is that there are different concepts of effectiveness. Fraser et al in reflecting on the replies to their survey hypothesised that auditors misinterpreted the question about effectiveness: *'It is possible that practioners are responding in terms of cost-effectiveness rather than effectiveness per se'* (Fraser, Hatherly and Lin). This illustrates two possibilities, that auditors are

concerned principally about the cost of obtaining their result, and/or that they misinterpreted the question in the postal questionnaire. The latter is a potential failing of all postal surveys. Interviews, although more costly for the researcher, allow the interpretation of questions to be monitored and controlled. Fraser et al claimed to have detected a difference in perception between the small firms and the rest. Small firms consider APs to be less effective than big firms do.

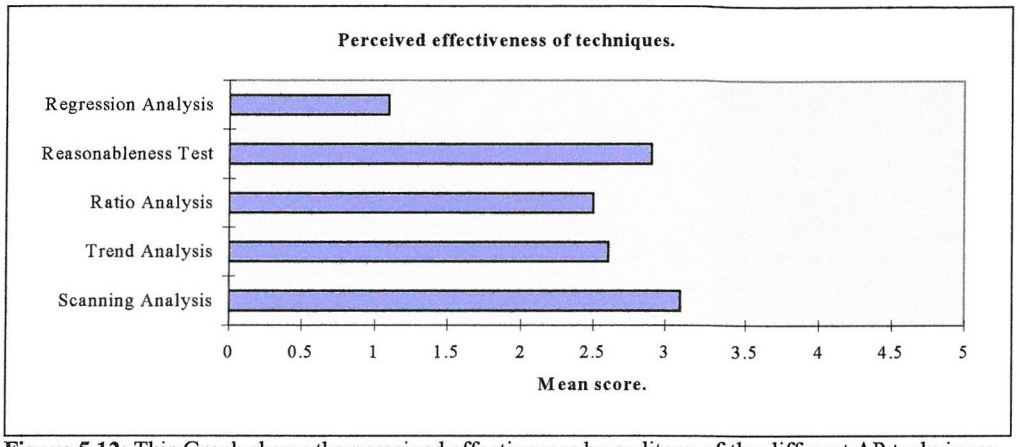


Figure 5.12: This Graph shows the perceived effectiveness by auditors of the different AP techniques on a scale of 1 to 5. Auditors appear to consider that simpler techniques that cannot be validated are best. Interestingly, such procedures have been shown by academics to be ineffective at detecting error. [Source: Fraser et al (1997)].

Fraser et al also asked about whether APs replace, or reduce, other substantive procedures. They found that if the respondents considered that the '*AP findings were favourable*' then other substantive were reduced or replaced.

5.6.1.4 Changes in the use of APs

Auditors were asked about the changes in the use of APs. The information from this question is summarised in Table 5.13.

	Increased use	Unchanged use	Decreased use.
Scanning Analysis	37%	63%	0%
Trend Analysis	61%	39%	0%
Ratio Analysis	67%	33%	0%
Reasonableness Test	57%	43%	0%
Regression Analysis	21%	78%	1%

Table 5.13: This shows that there has been more use of trend analysis, ratio analysis and reasonableness tests than other APs. Very few auditors thought that the use of any AP technique had decreased. [Source: Fraser et al].

The survey reported that one of the auditors' perceptions was that the use of APs had increased markedly over the last decade. The auditors surveyed by Fraser et al felt that since 1980 the usage of the procedures such as trend analysis, ratio analysis and reasonableness tests had increased, while a majority felt that scanning and regression methods had not. Very few auditors felt that the use of APs had decreased.

When the auditors were questioned about why they use APs, Fraser, Hatherley and Lin found that the auditors considered the competition between the firms to be a major factor. They also perceived APs to be effective at signalling error, APs initially signalling over 40% of all errors found by the auditor.

5.6.2 A Critique of the validity of this survey

There are a number of issues arising from Fraser et al's work. We know the non-response rate from the partners (46%), but although we know that the respondents were from 146 firms out of the 155 surveyed, we do not know the distribution of the non-response from the firms. If it was a small firm that might not affect the outcome, but if it involved one or more of the big six or the large firms that would be more serious. It is not clear how the individual partners were identified; for example were they a random sample from within each firm or were the questionnaires sent for distribution within each firm?

To test for potential bias from the non-responses, Fraser, Hatherley and Lin carried out a Mann-Whitney non-parametric test on the basis of firm size; they found that the two groups (responses versus non-responses) were not significantly different at the 5% level. They concluded that there was no bias arising from the high non-response rate.

Looking at the distribution of respondents there is some under-representation from the big six firms (by 14%) and from the small firms (by 21%). The large and medium sized firms are both over-represented: by 21% and by 5% respectively. When the information in the Fraser paper was recalculated to adjust for these under / over representations, the differences were not significant. The conclusions from Fraser are so strong that they are essentially unaffected by these under / over representations. If however we consider the problem of the non-response to the survey as a whole: 700 questionnaires sent out, 366 returned, we see that there could be some major changes to the reported results. Suppose that all those failing to respond to the survey did so because they do not use

APs. At this extreme, the percentages of those using APs would be reduced as shown in Table 5.14. The results at the planning and review stages are now more consistent with the usage reported by Ameen and Strawser (1994) and with the JMU findings reported by Garvey and Dietz (1997).

	Planning	Substantive	Review
Overall Average use of APs as reported by Fraser et al	60%	62%	86%
Overall Average if all those who did not respond did not use APs	31%	33%	45%

Table 5.14: This table shows the effect if the supposition that all those not responding in the survey did not use APs was true. [Source: Computed from information in Fraser et al (1997).]

Despite these shortcomings this survey provides a valuable picture of auditors' perceptions and use of APs. Because the use of APs is so high, the shortcomings of the survey do not alter the essential conclusions. APs are used extensively, simple techniques are the most common and auditors perceive that APs are effective.

5.7 The work of Mulligan and Inkster (1999)

The most recent published survey of the use of APs is that of Mulligan and Inkster (1999). Building on the work of Fraser Hatherly and Lin, it aimed, through a survey of audit partners, to investigate the extent of use of APs at the different audit stages, and the factors affecting their use. To address these objectives a questionnaire was sent to the offices of audit firms, who were asked to have the form completed by an audit partner. The sample is thus multistage and partly self-selecting. A total of 609 offices of 304 accountancy firms were mailed. The firms were stratified using the same criteria as Fraser et al. The sample can be summarised as in Table 5.15 below. Mulligan and Inkster tried to minimise non-response bias by a follow up mailing and checked for bias by comparing early and late responses.

	Sample				
Group	**Number of Firms**	**Number of Offices**	**Number of Questionnaires**	**Responses**	**Response rate**
Big six	6	59	150	46	31%
Large	34	200	200	92	46%
Medium	114	200	200	89	45%
Small	150	150	150	80	53%
Total	304	609	700	307	44%

Table 5.15: This table summarises the sample sizes for the Mulligan and Inkster (1999) survey.

Looking at Table 5.15, there are three issues that might bias the results. First, Mulligan and Inkster did not control who completed the questionnaire, so it is possible that the

questionnaires were handed to a member of staff with spare time and consequently the sample may not have been representative. Secondly, the low response rates from the big six audit firms, clearly not all offices replied. A third potential problem in interpreting the results is the under sampling of the small firms. Both this survey and that of Fraser, Hatherly and Lin (1997) the size of the sample of the small firms was smaller than it should have been, if the sample size was in proportion to the number of small firms, and they did not weight the results to reflect the under-sampling. However these small firms are perceived to be low users of APs, as demonstrated by the need for the ICAEW training courses: 'Tomorrow's Audit Today' that were run in 1998 to persuade these firms to adopt APs. Thus we have potential undercounting of the large firms who in other surveys are the heaviest users of APs and an under-sampling of small firms who are low users. In addition, there is always the possibility that the firms who did not reply failed to do so because they did not use APs at the planning and review phases where such procedures are required. With these caveats in mind, we now consider the key results from this survey, paying particular attention to the findings about the substantive use of APs.

There are several findings relevant to this thesis. First this survey confirms earlier work that simple techniques are used extensively at all stages. Secondly, APs are, in terms of the amount of time devoted to them, a major audit procedure. Thirdly, they are used to reduce other audit work. The review of this paper will concentrate on substantive APs, since they are the procedures of interest to us.

Mulligan and Inkster found that at the substantive phase, auditors tend to use the simple APs more often than the methods such as regression. For example, comparison, scanning and trend analysis, are used in at least 50% of audit applications, while regression is in only 15% of audits. Figure 5.16 shows this in more detail.

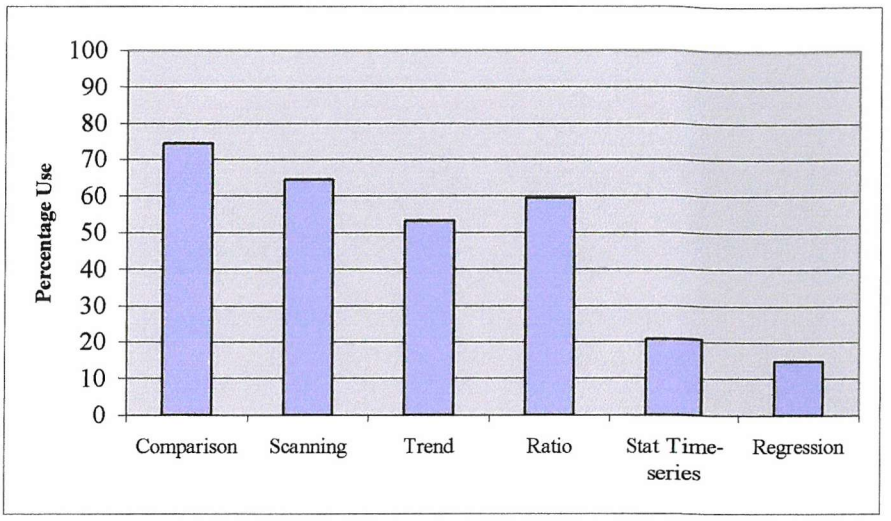


Figure 5.16: The usage of different AP techniques by the all firms at the substantive stage of an audit. [Source: Based on data from Mulligan and Inkster].

The respondents are indicating that in the audit of a set of financial statements they use more than one type of AP. That makes sense since in substantive testing the financial statements will be divided into smaller units, each of which, if it is material, will be subjected to a separate test. *'Regardless of the assessed levels of inherent and control risks, auditors should perform some substantive procedures for financial statement assertions of material account balances and transaction classes'* (SAS 300.8).

Figure 5.17 summarises the position by size of firm. It again shows that simple techniques such as comparison, scanning, trend analysis and ratio analysis dominate. It is clear from Figure 5.17 that firms of all sizes use the simple techniques extensively, and that the larger firms make more use than the smaller of the more 'sophisticated' procedures.

Because definitions of the techniques are not completely consistent with earlier research, some comparisons are difficult. Despite that caveat, the general finding is still that simple unstructured methods predominate. In all but one point, this confirms all the earlier surveys. The difference is in the application of regression type procedures, whereas Fraser Hatherly and Lin (1997) reported usage at the substantive of regression by the small medium and small sized firms of 17% and 15% respectively, much greater than Mulligan and Inkster's 6% for both medium and small firms.

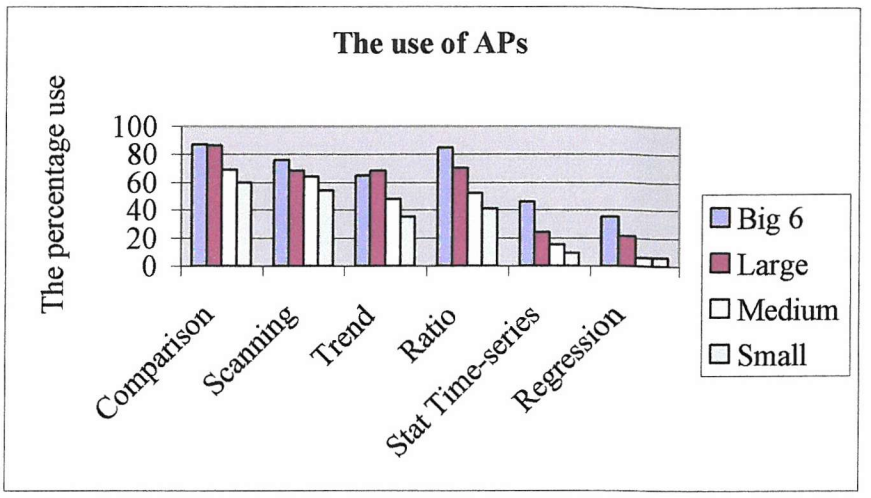


Figure 5.17: The graph shows the usage of the different categories of substantive AP, stratified by size of the firm. [Source: Based on data from Mulligan and Inkster.]

The next question is whether APs are used as a significant part of the audit. We can put some measure on this by using the findings about the relative time devoted to APs. Mulligan and Inkster (1999) reported that at the review phase of an audit, the use of APs was high. They found that: *'There were no statistically significant differences in the usage of APs at the review stage of the audit across all firms'*. They also found that the big firms were more likely to use APs at the planning stage. They surmised that: *'Given the involvement of small firms in accounts preparation, their more limited use of APs at the planning stage is expected'*. This gives some support to the JMU finding that APs were not documented as being done at the planning stage, and provides an explanation for that.

In Figure 5.18, we summarise the finding of Mulligan and Inkster concerning the amount of time devoted to APs at the substantive phase of an audit. They asked the Audit Partners to record the amount of time devoted to APs. The question was framed in bands of 0% to 20%, and comparable bands to 100%.

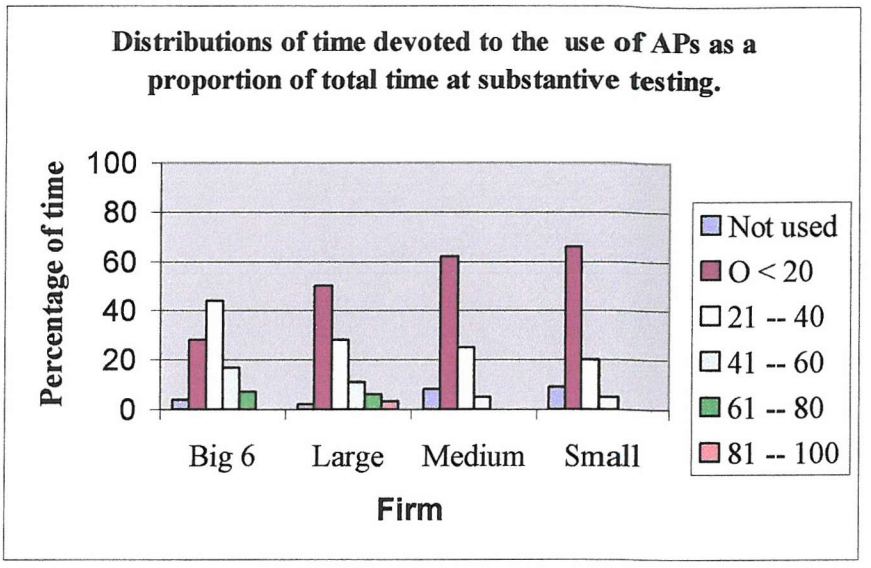


Figure 5.18: This shows the proportion of time devoted to APs at the substantive testing stage of an audit. Clearly there is an association between firm size and usage. For example for the small firm category two thirds of firms use APs for less than 20% of audit time. The equivalent figure for the big 6 firms is less than one third of the firms. [Source: Based on data from Mulligan and Inkster.]

At the substantive phase Mulligan and Inkster (1999) found that: '*Applying a Kruskal-Wallis non-parametric test for means, it was noted that the Big 6 and large firms were significantly more likely to use APs at the substantive stage*'. Inspection of Figure 5.18 confirms that this is so. If we consider the Big 6 and the small firms and look at the proportion of audits where the time devoted to substantive APs is more than 20%, we see that this is true of two-thirds of the audits conducted by Big 6 firms, while the equivalent figure for small firms is one quarter of audits. Overall, however it is clear that APs are an important substantive procedure.

The next question is whether there is any evidence for the view that APs replace or reduce other substantive procedures. The information from Mulligan and Inkster (1999) has been translated into Figure 5.19. Here the partners were asked to score 5 if they always replaced or reduced other tests of detail if they used an AP, down to a score of 1 if they never reduced the tests of detail. Some 63% scored 4 or more, and 94% score 3 or more, indicating that APs are used to reduce other substantive procedures.

Even if the concerns we raised on page 123 were valid, the important messages from this survey are so strong that they are still hold. This survey, like the others, clearly indicates that APs are a major evidence source at the substantive phase and that simple procedures dominate the type of AP used.

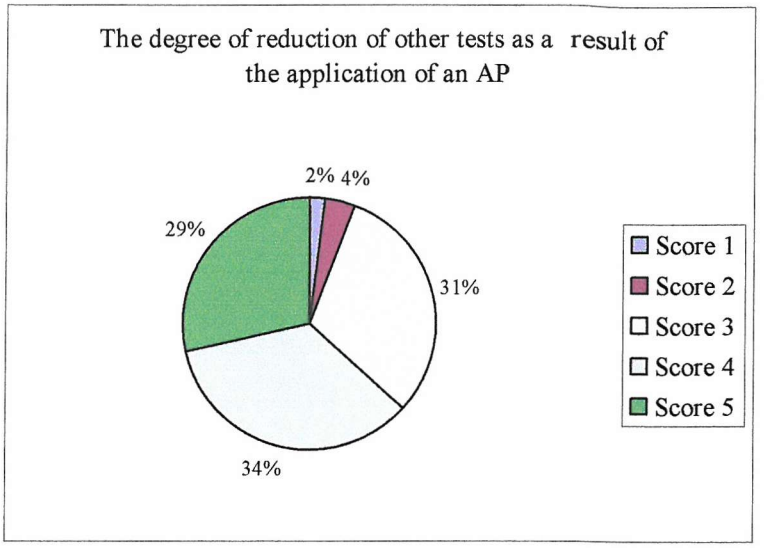


Figure 5.19: This shows the scores that auditors gave in response to the question 'how much do you reduce other audit test as result of the application of an AP?' A score of 5 indicates that test of detail are always reduced, while a score of 1 indicates that they never are. [Source: Based on data from Mulligan and Inkster.]

5.8 Summary of Finding from the Surveys

The surveys reviewed in this chapter have several themes in common. They are: APs are used a great deal, simple techniques predominate, a large (40%) proportion of all errors discovered are found when using APs, and APs are perceived by auditors to be cost effective.

5.8.1 APs are used at all stages of an audit

Firstly, the surveys reported that APs are used at all stages of the audit cycle: planning, substantive and review. However, the reported use at the planning and review audit phases conflicts with the observations of the JMU: *'Analytical Procedures should be applied at both the planning and overall review stages, but this rarely happens'* (Garvey and Dietz (1997)). If one looks at the survey of Fraser, Hatherly and Lin (1997), they found that at the planning stage there was an association between the use of APs and firm size, with the larger firms making more use of APs. There was a marked difference between the big 6 firms and the rest, although even the smallest firms appear to use APs at the planning stage in over 40% of their audits. The use of APs at the review stage was consistently high. The survey by Ameen and Strawser (1994) signalled that APs were used for some 40% of the total audit hours, providing evidence that is consistent with the Fraser survey.

A possible explanation for these differences over the use of APs at planning and review is that the firms define certain procedures as being APs, which the JMU did not. Alternatively, the explanations for the discrepancy may be poor documentation of the APs, which if correct, needs to be addressed. The rationale for the investigation rule needs to be recorded so that the AP can be replicated, either in management review or for quality control purposes. This is very important for simple APs where there frequently will not be a recognized model to use as the basis for re-performance. All procedures should have the ability to be replicated as one of their attributes. *'Record the procedures carried out and document matters which are important to support conclusions reached or statements provided in the report'* (From Appendix 8 of Practice Note 11: The Audit of Charities).

A third possible explanation is that, while the JMU is looking at all auditors, the other surveys are biased towards the larger firms. We have already seen that there seems to be a link between firm size and the use of APs. Although small firms were not excluded from the surveys, they were under-represented. As further evidence that the smaller firms make relatively little use of APs, we note the recent emphasis of the ICAEW professional training program. For example in 1998 the Audit Faculty of the Institute of Chartered Accountants of England and Wales ran a series of seminars for their members: - 'Tomorrow's Audit Today'. It was targeted at the small firms and the message was that the audit should be economic (low cost) and add value for the client. A key theme[1] was the *'elimination of transaction testing and their replacement by Analytical Procedures'*. Thus, one possible argument is that small firms who tend not to use APs were under-represented in the surveys and consequently the use of APs may have been exaggerated. Fraser's (1997) observation that *' because smaller firms use APs less than larger firms they will find less error relative to their other audit work'* is consistent with this line of argument.

5.8.2 Simple APs are performed at all stages of an audit

A second theme from the surveys is the predominance of simple APs such as scanning, trend analysis and ratio analysis. This is true at all stages of the audit cycle: planning, substantive and review. One would expect simple APs to be used at planning and

[1] Contained in the official session notes of Tomorrows Audit Today.

review. Indeed the Auditing Standards imply that this should be the case. For example when discussing planning, SAS 410 says, *'Analytical procedures at this stage* (planning) *are usually based on interim financial information, budgets and management accounts'* (SAS 410 paragraph 10 in the section on planning procedures). In contrast when discussing substantive procedures the SAS goes into far more detail. For example, it makes the point: *'When intending to apply analytical procedures as substantive procedures, auditors consider a number of factors such as: the degree to which information can be disaggregated, for example analytical procedures may be more effective when applied to financial information on individual sections of an operation or to financial statements of components of a diversified entity, than when applied to financial information relating to the entity as a whole'* (SAS 410, paragraph 13 in the section on substantive procedures). Disaggregated data tend to be associated with substantive APs while higher-level data are usually sufficient for carrying out simple APs at the planning phase of the audit. Loebbecke and Steinbart (1987) also made the association of simple APs with annual data: '*There is significant evidence that preliminary analytical procedures, i.e. procedures using annual financial statement data, are being used by many auditors'*. They then imply that such procedures are used to gather substantive evidence: *'Given the increasing pressures to minimize audit costs and move towards more structure in audit in the planning process, we believe that some, perhaps many, auditors may be using preliminary APs not only as attention directing devices, but as substantive evidence in support of specific financial statement assertions as well'*. Thus the second important piece of information from the surveys is that auditors gather substantive evidence using simple APs, possibly using high-level data.

5.8.3 APs account for about 40% of all errors detected

A third theme is that auditors report that an AP first signals over 40% of all errors discovered during an audit. In their survey, Biggs and Wild (1984) found evidence that APs initially signalled 42% of all errors, and subsequent surveys have confirmed that figure. Tabor and Willis (1985) and Ameen and Strawser (1994) reported that about 40% of all audit time was devoted to APs. The two findings, the proportion of errors discovered by APs and audit time devoted to APs, are consistent with each other.

There is a piece of research that runs contrary to the evidence about the level of error discovered by APs. It comes from Bell and Knechel (1994) who surveyed a small number of partners responsible for the audit of Property and Casualty Insurers in the USA. One of the issues that came out of the survey was that of all procedures used in an audit, APs produce 15.8% of all error signals (1.9% at the planning phase of an audit, 10.2% at the substantive phase and 3.7% at the review phase). This finding is very different to that from the other surveys reported in this Chapter. It was from a very detailed analysis of a small survey targeted at analysing all errors found during the audit process, whereas the other surveys were looking at APs in particular. More importantly it relates to a very specialised part of the market: of property and casualty insurers in the USA. Even with that caveat and the small size of the sample (28 audit engagements) it raises the question of whether the auditors' responses to the other surveys were exaggerated. Those surveys were dealing with auditors' perceptions; Bell and Knechel (1994) were analysing documented facts.

The auditors surveyed by Fraser et al (1997) felt that the usage of APs had increased markedly over the decade prior to the survey, particularly procedures such as trend analysis, ratio analysis and reasonableness tests. If that were the case one could reasonably expect that there would have been an increase in the proportion of errors discovered by the use of APs. An alternative argument might be that these simple APs are not so effective as the more complex APs in identifying error, and it is because the use of the former had increased there had not been a corresponding increase in the number of errors being found. Unless this is the case, the high level of usage reported by Daroca and Holder (1985), nearly as high as the more recent surveys, leaves the suspicion that their results were biased. It may well be that Hylas and Ashton (1982), when they reported that 27% of all errors were signalled by APs, were nearer the truth for the early 1980's. That would be consistent with the argument that the early surveys were biased and overstated the effectiveness of APs in discovering error.

5.8.4 Do auditors always apply APs correctly?

In Chapter 4 we raised a potential weakness for the application of APs that was identified in the work of Humphrey and Moizer (1990) and of Garvey and Dietz (1997). This concerned the possible imperfect understanding by auditors of the different techniques they use, and consequent risk that APs may not always be properly

applied. Whether this is an important issue for APs depended on the extent and nature of the use of APs. We now can see that APs are used extensively and that simple APs predominate. What the surveys have not considered, or at least have not reported, are the procedural issues. These include aspects of both investigation and outcome rules. For example, within the investigation rules, what is the quality of data used and what checks are there of such data. Also what quality control measures are used for the corroboration of the APs and how they are documented? Since APs are often informal in the sense that they are not mathematically based, the rationale for the AP should be recorded so that the AP can be replicated. There is a requirement for this; the mandatory part of SAS 230 on working papers states that: '*Working papers should record the auditors' planning, the nature, timing and extent of the audit procedures performed, and the conclusions drawn from the audit evidence obtained*' (SAS 230.2). Paragraph 7 expands on this: '*The extent of working papers is a matter of professional judgment since it is neither necessary nor practical to document every matter auditors consider. Auditors base their judgment as to the extent of working papers upon what would be necessary to provide an experienced auditor, with no previous connection with the audit, with an understanding of the work performed and the basis of the decisions taken*' (SAS 230: Working papers – author's underlining). This part of the SAS implies that unless the auditor records a sufficient description of the rationale of the AP to allow the reviewer (the experienced auditor in the quotation) to follow the calculations then it will not be possible to form a judgement on the adequacy of the audit for that specific audit assertion and thus meet the requirements of this SAS.

Another key stage of an AP, the outcome rule, has not been addressed by the surveys. It is the process of identifying and resolving significant differences between the expectation value and the book-value in the financial statements. An important part of that outcome rule is obtaining and testing an explanation for the difference.

5.8.5 Are APs effective?

Despite the problems of non-response and some apparent contradictions (e.g. between most of the surveys and the work of Bell and Knechel (1994) – discovered error is 40% vs. 16% of all errors found), the surveys have provided a valuable snapshot of the extent of use of APs. There can be no doubt that APs are a major source of evidence: APs are used extensively, simple techniques are the most common and auditors

perceive that APs are effective or, possibly, cost effective. This is particularly true at the substantive phase of the audit, our main concern in this thesis.

There are some of the issues about the quality of the AP work that have not been considered, the next chapter will address those in a new survey carried out for this thesis. For example, it will consider whether the data is reviewed as required in the Standards, what quality control work is done on the data used as a basis for the APs or on implementation of the investigation-rule. Another issue explored is whether auditors actively consider the implications of the explanation-effect.

The Literature Review will then be completed in Chapter 7, which considers some experiments into the effectiveness of APs.

Chapter 6: A new survey of the use of APs in the UK.

6.1 Introduction

This chapter describes a survey undertaken as part of the research for this thesis. It addresses some of the issues not dealt with in the surveys described in Chapter 5. My survey was done by face-to-face interviews with the Technical Partners of the largest sixteen audit firms in the UK. It was carried out over the period 1994 to 1995. The survey topics included the extent of use of APs and the nature of the investigation and outcome rules.

6.2 The aim of our survey

One of the disadvantages of postal surveys is the possible misinterpretation of a survey question. An example occurred in the survey by Fraser, Hatherley and Lin (1997), which we described in section 5.6. Auditors interpreted a question about whether APs were effective to be, were those APs cost-effective? To minimize such risks, to probe issues and ensure a high response rate we chose to use face-to-face interviews, based around a questionnaire, a copy which is in Appendix A.

It was recognised that it would be a time consuming exercise. To minimize the time involved, while obtaining the maximum information, we decided to target the sixteen largest audit firms in the UK. They are the firms that have the economic size to develop and maintain audit methodology and audit manuals and to maintain those.

The aim of the survey was to expand on the information gathered in the surveys described in Chapter 5. The survey first looked at the definition of an AP, then at the extent of use of APs, which procedures are currently used and at what stages of the audit. Also considered was the data used, data quality and what checks there are of such data form an important part of the investigation rule. Information was then gathered on that, and the quality control measures carried out to corroborate the APs and how this work is documented.

A key part of the outcome-rule of an AP is the process of resolving the difference between the expectation-value and the book-value in the financial statements. So the next questions

addressed the procedures the firms have in place for obtaining explanations for that difference and the steps taken to address the explanation effect problem we described in Chapter 4 (section 4.10).

We wanted to establish the role APs play as substantive procedures. So to that end we sought to establish the extent to which APs replace other forms of substantive tests, such as sampling.

6.2.1 Who formed the target population of the survey?

Within the major audit firms, technical policy is developed and disseminated by the Technical Partners. They are well placed to have an overview of the extent of use of APs across the firm. This group was the target population.

Although the Auditing Standards are common to all audits, they represent the minimum set of criteria that the auditor must meet. The firms each puts into practice its own audit methodology, which must comply with the Auditing Standards. An analogy is that the Standards form the skeleton of the audit process; the audit manuals put the flesh on the Auditing Standards to form a whole audit process that is unique to a particular Audit Firm. The surveys described in Chapter 5 investigated the staff charged with implementing the policy and who face the practical difficulties of using APs. The new survey represents the views of a different group within the profession: the policy makers rather than those implementing the policies who were the target of the earlier surveys. Any difference between the perceptions of these two groups is of interest.

A brief explanation of the profile of the private sector audit firms will put the survey strategy into context. The population of audit firms can be divided into two distinct groups: those who develop their own methodologies and those who import methodologies. In addition there are wide variations in the size of the audit firms. The private sector firms can be allocated into three strata: the big six audit firms, the next ten audit firms and the remainder of the audit firms. There are thousands of these small firms, varying in size from several partners to single partner practices. Broadly the first two strata are the group who

develop their own methodologies, while the smaller firms in stratum three normally import their methodologies.

Six big firms in the private sector (UK) dominated the audit of private sector companies. As at August 1995 they were:

Coopers and Lybrand	Ernst and Young	Arthur Andersons
KPMG Peat Marwick	Price Waterhouse	Touche Ross

Since the survey, there have been changes, the most significant being the merger of Price Waterhouse and Coopers and Lybrand to become PricewaterhouseCoopers.

The relative size of these firms can be put into perspective by considering the income from audit fees of the firms. Audit fees rather than income are used since income reflects the wider activities of the firms, rather than simply their audit work. For the large firms, this did not make a difference in the selection of the firms, since most of the firms collect substantial fees for non-audit work. But in the second category of firms it did make some differences.

In 1994 when the survey was being planned, the income from audit fees of the largest of the big six firms (Coopers and Lybrand) was £253m, and that of the smallest (Touche Ross) was £122m. At that time, the largest firm of the next group of firms was Grant Thornton whose income was about one third of that of Touche Ross. Now, in 2001, the largest firm in stratum two is BDO Stoy Hayward, but the same proportionality still applies for the differential between what are now the big five and the next group of audit firms. All these firms in stratum two are large and are important players in the development of audit policy. As at August 1995, the time of the survey, they were: -

Grant Thornton	Moore Stephens
BDO Stoy Hayward	Robson Rhodes
Pannell Kerr Forster	Neville Russell
Clarke Whitehill	Moores Rowland
Kidsons Impey	Baker Tilly

To put the relative size of the firms in this second tier group into context, Baker Tilley was the firm with the smallest audit fee income: £18m. In contrast, the average of the next 34

firms in the top 50 audit firms was £3.9m. The table in Table 6.1 summarises the difference in income from audit fees between the firms.

	Audit Fee Income (£m)	Average (£m)	Audit Fee Income as % of total income of the top 50 firms
Big Six Audit Firms	£1048m	£174.7m	72%
The Next Ten	£271m	£27.1m	19%
The top 34 of the smaller firms	£131m	£3.9m	9%
Total of top 50 Firms	£1450m		

Table 6.1: The distribution of fee income from audit services in 1994/5, the year of the survey. [Source: Computed from data published by ICAEW.]

Only the larger firms can afford to support a technical team to develop audit policy, training and to produce the associated audit manuals.

6.2.2 The sampling methodology

In the three strata described above all the firms in strata (1) and (2) were included in the survey. The dominance of the larger firms in terms of the fee income, and their role in research and the development of methodology were the main reasons for this decision. The smaller firms frequently do not have the resources to develop audit policy and manuals. Those small audit firms were not surveyed in detail since there was a prior expectation that they import the audit procedures and manuals from the larger firms. We confirmed this assumption by talking to a number of small firms. Some of the small firms enter a formal association with a larger firm (e.g. KPMG) who supply the audit methodology, often in the form of training and audit manuals. On the other hand, we found one instance of a group of small firms who combined to form a co-operative to develop their own audit manuals.

The use of APs by the small firms we talked to was consistent with that of similar firms in the Fraser survey. That is, those small firms use APs less than the larger firms and do not consider them to be so effective.

6.2.3 The data collection methods

The method of data collection for the survey was by interview with the technical partner[1] of each audit firm. The interview was structured around a questionnaire, which ensured that the same basic data were collected from each person in the survey. The face-to-face interviews allowed the opportunity to clarify any potential misunderstandings or misinterpretations of the question or reply. The face-to-face interviews made it possible to explore the different approaches of the firms and to identify new developments. Within the confines of confidentiality, the interviews provided the firms with direct feedback and allowed a two-way dialogue. It addressed key failings of many postal surveys: the possible misinterpretation of the questions and the difficulty of exploring in depth issues behind the basic question. A postal survey would not have allowed the building of confidence that enabled much in depth information to be obtained. My survey was thus more qualitative than most of the other surveys.

6.3 Results of the Survey

6.3.1 The types of AP used at the planning, substantive and review phases of an audit

Figure 6.2 shows the relative use of different types of APs at the three phases of an audit.

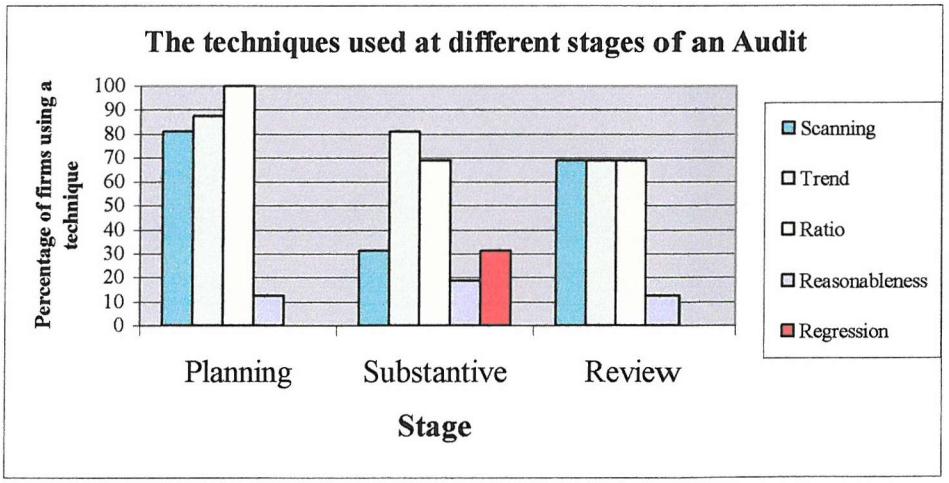


Figure 6.2: The use of the different types of AP at different stages of an audit. (Source: This thesis).

[1] In one case the AP strategy was the responsibility of a senior audit manager. In that instance the manager was interviewed rather than the technical partner.

In this chart, the height of the bar represents the percentage of firms using a particular type of AP at the relevant stage of the audit. The auditor will frequently divide the financial statements into many discrete areas, each of which may be subjected to an AP. Thus an auditor may use all the different types of AP on a set of financial statements. In discussion they asserted that they choose what is the most appropriate for a particular attestation.

It can be seen that firms use a mixture of techniques at any stage, but from an inspection of the data in the chart at Figure 6.2 it is clear that the simple procedures are the most commonly used by auditors. This is true at all stages of an audit.

At the substantive phase, the full range of procedures is used, but the simple ones predominate. At the planning and review stages this extensive use of simple methods is consistent with the earlier surveys. For example about 80% of firms use a scanning type of procedure at the planning stage. All firms use ratio analysis at the planning stage. My survey confirms the supposition made in section 5.6.1.1 when considering the use of reasonableness methods at the review stage. There was a problem of different definitions of reasonableness procedures between those devising the questionnaires and those answering them: auditors do not make much use of such methods at the review phase of an audit.

The simple APs are used to identify areas of potential audit risk and to assist in planning the nature, timing and extent of other audit procedures. *'Analytical procedures at this stage* (planning) *are usually based on interim financial information, budgets and management accounts'* (SAS 410). Although this wording does not, in itself, instruct the auditor to use only simple procedures, the SAS is more specific about the form of the investigation rule for the AP when considering substantive APs. For example, SAS 410 says auditors should consider: *'the degree to which information can be disaggregated, for example analytical procedures may be more effective when applied to financial information on individual sections of an operation or to financial statements of components of a diversified entity, than when applied to financial information relating to the entity as a whole'* (SAS 410). An implication is that high-level data are appropriate, or at least adequate, for planning APs. The SAS appears to imply that, other than at the substantive audit phase, simple APs

using high-level data will be sufficient. That does not explain why the simple procedures predominated at the substantive stage as well. We will return to this at the end of the chapter.

The issue of which APs are used was useful for two reasons. First it provides evidence about the types of procedures the firms expect to be used and secondly it provides a benchmark with the other surveys we discussed in Chapter 5. This question covered the same area as the other surveys, and the extent to which it agrees with them provides some credence for the answers to questions that were not raised those surveys.

6.3.2 Sources of the data used in APs and the checking of it

The quality of evidence from an AP will depend on many factors, particularly the quality of the data on which it is based. The Auditing Standards observe that: *'The reliability of the information used in analytical procedures is likely to be enhanced if it comes from sources independent of, rather than internal to, the entity. If the information is produced internally, its reliability is enhanced if it is produced independently of the accounting system or there are adequate controls over its preparation. The necessity for evidence on the reliability of such information depends on the results of other audit procedures and on the importance of the results of analytical procedures as a basis for the auditors' opinion'* (SAS 410). The SAS is arguing that if the AP is an important source of audit evidence for a particular assertion, then it is necessary to obtain evidence on the reliability of the data used in the AP. The Auditing Standards suggest that one avenue of investigation of the data will be to assess what information the audited entity possesses and to consider its reliability: *'Auditors usually enquire of management as to the availability and reliability of information needed to apply analytical procedures and the results of any such procedures performed by the entity. It may be efficient to use analytical data prepared by the entity, provided the auditors are satisfied that such data is properly prepared'* (SAS410, paragraph 12).

The survey considered the extent to which auditors check the reliability of the data they use as part of the AP process. Such data is a key part of the investigation-rule. In the interviews with the Technical Partners there were several questions about the data used in

the APs. All the firms make extensive use of data from both internal and external sources. On the question of investigating the data the result was interesting. About one-third of the firms (6 Firms) make no effort to confirm the quality of the data. The details are shown in Figure 6.3.

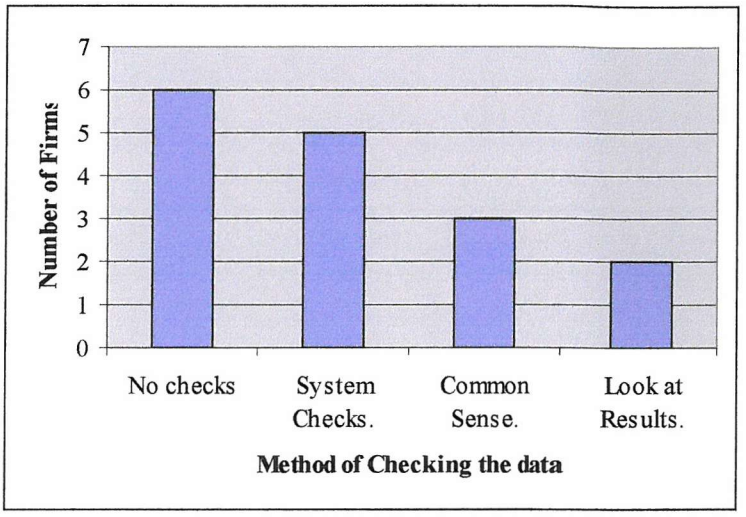


Figure 6.3: The relative amounts of audit work done by auditors into the quality of data used to support their substantive APs. [Source: Survey for this thesis.]

We first consider those firms who do not investigate data quality before the application of the AP. They were split into those who were quite candid and said that they did no checks of the data (6 Firms) and those who said that if the result of the AP was right the data must have been of good quality (2 Firms). When pressed to define 'the answer is right' they said that if the value that they calculated, the expectation value, was close to the book value in the financial statements being audited, then that would be 'the right answer'. This is, implicitly, the outcome-rule: $|\hat{Y}_T - Y_B| \leq C$; where C is the critical value they use in their outcome-rule. In discussion the partners either set C at the overall materiality, or declined to define C.

The firms, who did some checking of the data, split into two sub-groups, those who did systems checks or some form of data confirmation (5 Firms) and those who instructed their staff to use their common sense (2 Firms). Thus only one third of the audit firms positively

claim to do checks that the data are reliable, and half of the firms do not quality assure the data.

6.3.3 Quality of the explanations if $|\hat{Y}_T - Y_B| > C$

The SAS on APs requires: '*When significant fluctuations or unexpected relationships are identified that are inconsistent with other relevant information or that deviate from predicted patterns, auditors should investigate and obtain adequate explanations and appropriate corroborative evidence' (SAS 410.4).* The debate is about how the explanations are obtained and what is sufficient evidence. The survey considered the procedures the audit firms set as their policy for obtaining explanations of such differences when $|\hat{Y}_T - Y_B| > C$. The quality of such explanations was perceived by three firms to be a significant weakness of the AP process. They were worried about the problem of obtaining adequate and reasoned explanations for the differences between the predicted and book values.

As a result of the question in the survey and follow-up discussions with the firms we found that five of the firms have no requirement to obtain explanations when $|\hat{Y}_T - Y_B| > C$, nor do they have any training on the subject. Six of the firms had no requirement in their audit manuals to obtain and document explanations but did include the subject in their training. Five of the firms had a requirement to obtain explanations when $|\hat{Y}_T - Y_B| > C$ and also covered the subject in their training program. Thus eleven of the firms do recognise the subject, but of these eleven, only three had in place a strategy to ensure that explanations and alternative explanations are considered before seeking clarification from management.

It is worrying that few firms, only three, adopt the strategy described in section 4.10 to counter the explanation effect problem. The firms do not appear to have learnt from the work in the social sciences (R Libby (1981 and 1985) and Anderson and Sechler (1986)) about the need to ground alternative explanations. There should be a rigorous approach to obtaining and checking explanations for the difference between predicted and audited figures.

An additional piece of information emerged from the discussions. It was found that auditors rarely consider an explicit figure for the precision. Thus the decisions on the closeness take no account of the precision associated with the expectation values, and they are, in fact, treated as if they were exact values. As argued earlier such a requirement is not explicit in the UK Auditing Standards although it is in the USA Standards. The impact of this is discussed in the next chapter, which considers the performance of particular APs.

6.3.4 The motivation to use APs and the necessary conditions to use APs

We then addressed the issue of what motivates the auditor to use APs. At the substantive stage the primary motivation was cost, cited by eleven of those questioned. However, half of the Partners considered APs to be good at detecting error. Two of the firms believed that APs are a quality procedure that delivers high quality evidence. Many of those using APs at the substantive stage emphasised that it is necessary to understand the business to implement a successful AP. That is logical; it is necessary to understand the relationships between different sets of data to build an expectation-value. The objective is to check whether a value in the financial statements contains material error or is free of it.

The need to understand the business in order to construct an effective substantive AP highlights a contrast in the purpose of APs at the different phases of an audit. At the planning stage the purpose of the AP is to explore the relationships between different data in order to help build a picture of the inter actions within and between the different systems or elements that make up the financial statements. The Partners, eight of whom said that APs assist in understanding the business, confirmed this. At the substantive stage the purpose is to use the knowledge of the entity and its systems to predict a value.

6.3.5 The extent to which APs replaces other forms of substantive tests

The concept of combining audit evidence has been implemented in a variety of ways. Some firms have adopted a form of Audit Risk Model, other firms have devised a scoring system where the scores from different levels of evidence are predetermined, and a third method is through tree diagrams. The Audit Risk Model was discussed in Chapter 2.

One of the questions in the survey asked: 'How is evidence from APs combined with other substantive evidence?' This enabled us to establish, in discussion, the extent to which other procedures (e.g. sampling) are replaced by APs. Figure 6.4 summarises the replies.

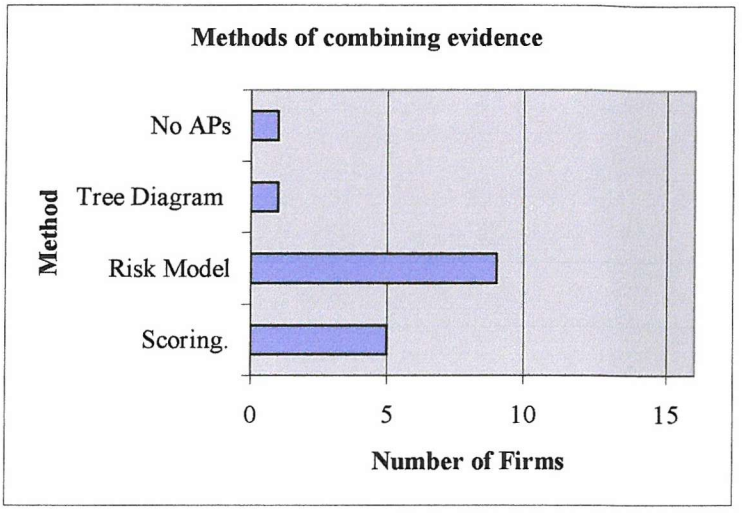
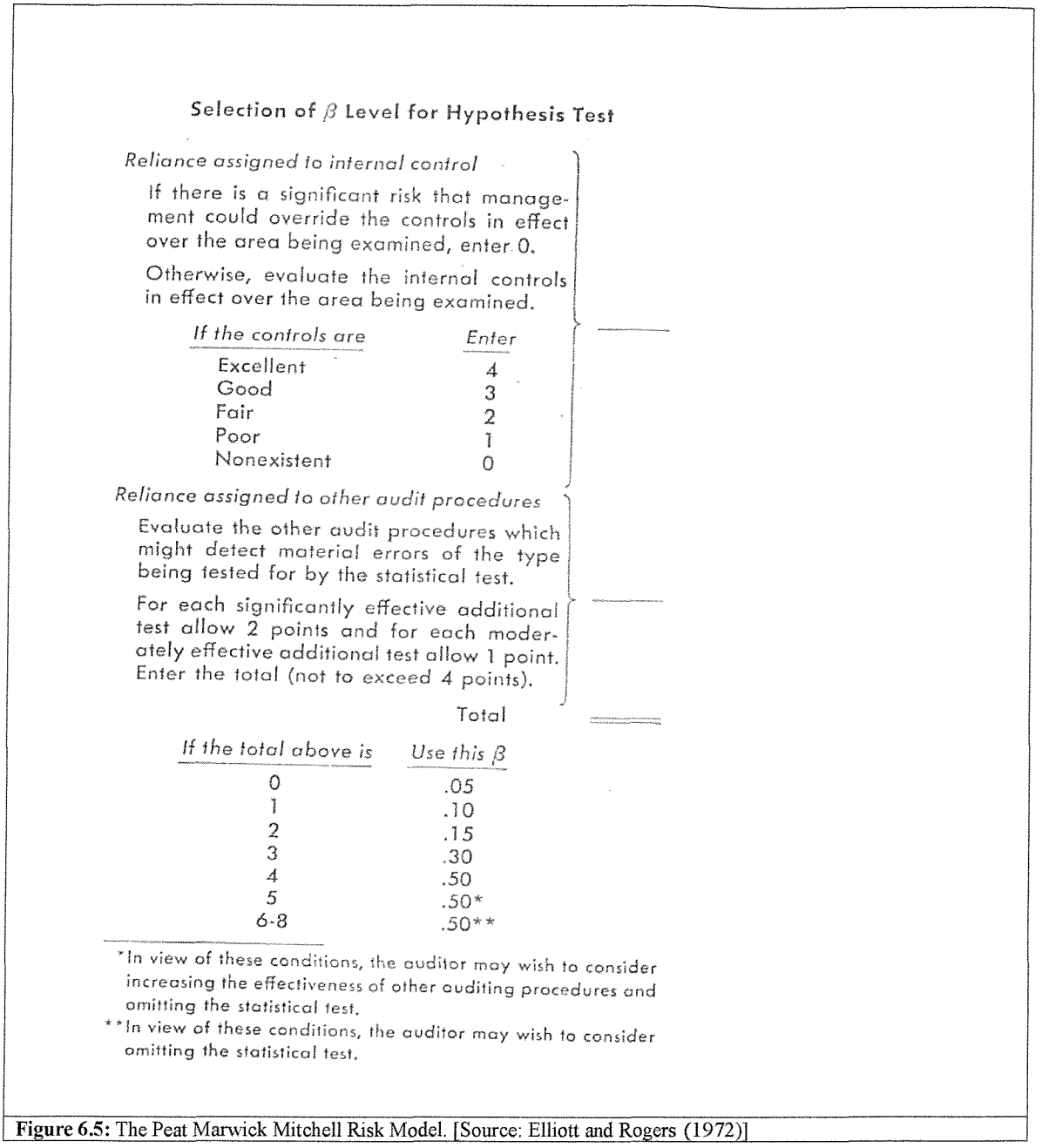


Figure 6.4: Ways in which the firms combine the substantive audit evidence. [Source: Survey for this thesis.]

All the firms, except for one, use APs at the substantive phase of the audit. Those using APs, combine the signals from APs and other audit work using a form of the audit risk model. It may be the common form shown discussed in Chapter 2, a scoring system or a decision tree.

The aim of a scoring system is to assist the auditor in deciding the level of substantive sampling. A typical scoring system was that adopted by Peat Marwick Mitchell (now part of KPMG) and is shown in Figure 6.5. This scoring model reflects the requirement for some level of substantive sampling in the Audit Manuals of the 1970s. The diagram is taken from a paper by Elliott and Roger (1972), and it was still present in the 1979 Audit Manual of Peat Marwick Mitchell. The scoring system worked by the auditor assigning scores, which in their judgement reflected the strengths of the controls and the effectiveness of substantive procedures other than sampling. The result of applying the scoring system would be a measure of the amount of evidence required from a substantive sample.



Selection of β Level for Hypothesis Test

Reliance assigned to internal control

If there is a significant risk that management could override the controls in effect over the area being examined, enter 0.

Otherwise, evaluate the internal controls in effect over the area being examined.

If the controls are	*Enter*
Excellent	4
Good	3
Fair	2
Poor	1
Nonexistent	0

Reliance assigned to other audit procedures

Evaluate the other audit procedures which might detect material errors of the type being tested for by the statistical test.

For each significantly effective additional test allow 2 points and for each moderately effective additional test allow 1 point. Enter the total (not to exceed 4 points).

Total

If the total above is	*Use this β*
0	.05
1	.10
2	.15
3	.30
4	.50
5	.50*
6-8	.50**

*In view of these conditions, the auditor may wish to consider increasing the effectiveness of other auditing procedures and omitting the statistical test.

**In view of these conditions, the auditor may wish to consider omitting the statistical test.

Figure 6.5: The Peat Marwick Mitchell Risk Model. [Source: Elliott and Rogers (1972)]

Using the scoring system shown in Figure 6.5, suppose that the auditor after reviewing and testing the systems of financial controls decides that the systems are 'Good': that would score 3. If there were two types of substantive test, APs and the sampling and the auditor decided that the reliance on APs was 'effective' at detecting material error, then that might score 2. The total score would be 5. We can see from Figure 6.5 that this would mean that

the statistical sample size would be calculated to achieve a ß-risk of 0.5, the equivalent of a 50% Confidence Level for the sample. The auditor does have the option to override the requirement for a substantive sample. An alternative example would be if judgements were that the systems were good, again a score of 3, and that no other substantive tests where done (a score of 0), and then the total score would be 3. For the scoring system in figure 6.5 this would imply a ß-risk of 0.3, the equivalent of a 70% Confidence Level for the substantive sample.

One firm had adopted another version of the Audit Risk Model: it is a form of a tree diagram that assists the auditor's decisions. The diagram is shown in figure 6.6.

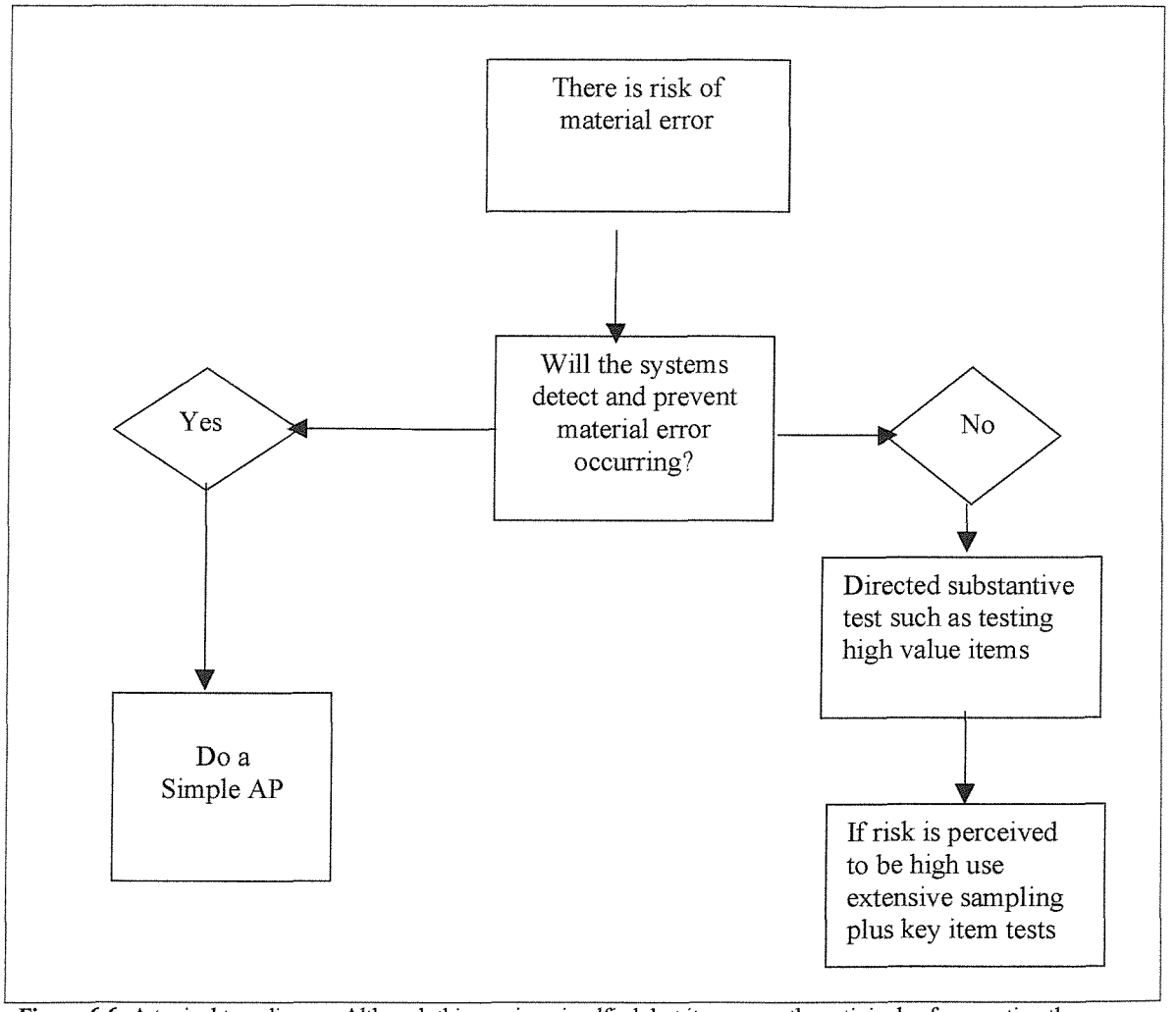


Figure 6.6: A typical tree diagam. Although this one is a simplfied, but it conveys the ratioinale of converting the auditors decisions about the systems and other work into a strategy for the type, and extent, of sustantive tests. [Source: Survey for this thesis.]

A description of that firm's process of bringing the evidence together is provided, and although it is more formalised than some, it conveys the rationale of converting the auditors decisions about the systems and other work into a strategy for the type, and extent, of sustantive tests. It is the system, for example, adopted by the NAO.

This tree-diagram is typical of the way the profession has moved in the late 1990's. It focuses the auditor's attention on the perceived risks and then directs the auditor towards the type of substantive evidence required. It was clear from the interview that the expectation is that this will usually be some simple AP.

To use the tree-diagram, first the auditor makes a judgement about the risks to the financial statements, (i.e. risks that material error might occur) and also about the risks of the systems not detecting material error. If the systems are judged to be good and the risks to the auditor are low, then simple APs are used to provide the required substantive evidence. This scenario is shown on the left hand branch of the tree diagram.

If the systems are judged not to be satisfactory, or there is some doubt about the systems, or there is a perceived risk to the part of the financial statements being audited; then the approach is not to adopt APs. Instead, the auditor should perform directed substantive tests that address the perceived risk. This might take the form of testing a large sample of high-value transactions or those coming from a particular system, as indicted on the right-hand branch of the decision-tree.

In effect it is a decision path that the auditor follows and is not a formal combination of evidence. The decision tree provides the alternative avenues for the auditor to follow. It provides the path, once the auditor has judged what the risks are and adequacy of the systems, to prevent material error. It provides guidance where it is needed. With regard to APs, in this particular system, the choice is whether to use them or not. If the risk of material error is perceived to be low, then a simple AP is sufficient, but otherwise some other audit procedure has to be adopted. The cynic might say that the AP is adopted as a way of satisfying the SAS 300 requirement that some substantive work must be done: *'Regardless of the assessed levels of inherent and control risks, auditors should perform*

some substantive procedures for financial statement assertions of material account balances and transaction classes' (SAS 300.8).

Thus in this decision tree approach the APs adopted are simple ones, and modelling type APs such as reasonableness tests or regression are not used.

6.3.6 Changes in the use of APs over the five years to 1995

We also asked the Technical Partners about how they saw the extent of changes in the use of substantive APs. All felt there had been some increase in the use of APs over the last five years. It was the simple procedures that were considered to have increased most. This information is shown in Table 6.7; it reflects the proportion of respondents who thought that the use of APs had increased, decreased or not changed.

	Increased use	Unchanged use	Decreased Use
Scanning Analysis	37%	63%	0%
Trend Analysis	63%	37%	0%
Ratio Analysis	69%	31%	0%
Reasonableness Tests	56%	44%	0%
Regression Analysis	19%	81%	0%

Table 6.7: The perceptions about the changes in the use of APs. [Source: Survey for this thesis.]

Although the majority of respondents considered that the use of simple procedures such as trend or ratio analysis had increased, most thought that the use of more 'complex' procedures such as regression analysis had not changed over the last five years. When asked a similar question about the future use of APs, only 3 firms thought there would be a further increase in the use of APs. Although the firms think APs have reached the limit of their use, none thought that there would be a decrease. APs are considered to be an essential procedure for the future.

Technical Partners were asked for their opinions of why their firms use APs and then why other firms use APs. Table 6.8 below summarises the replies to the interviews. The firms were asked which of these were important: they could choose all or none. Although most gave the same answer for both questions, it was interesting that on some issues this prompted some differences in the replies. For example two firms considered that the cost of APs was not an issue for them, but that it was for other firms.

Reasons given for the use of APs	Why do other firms use APs?	Why does your firm use APs?
1. APs are perceived to be effective at detecting error.	6	6
2. They help with the understanding of the business.	3	3
3. APs are inexpensive.	16	14
4. Quality process.	2	3
5. Overcomes mechanical audit.	2	2

Table 6.8: Auditor's perceptions of why APs are used. The data record the number of firms. [Source: Survey for this thesis.]

APs are considered to be an inexpensive procedure to implement (all of the firms) and, a substantial number thought them also to be effective at detecting error (6 of the firms). It was quite clear when discussing this with the firms that they interpret effective to be cost-effective. This is consistent with the interpretations by Fraser, Hatherly and Lin (1997), as discussed in Chapter 5, of the opinions on effectiveness in their postal survey.

Points 4 and 5 of Table 6.8 might appear to be different versions of the same attribute. APs are considered to be a quality procedure and are vehicles for overcoming mechanical audit. However all is not as it seems. An example of mechanical audit was given as sampling, and the quality APs were those that allowed the auditor freedom, *'not to be shackled by the dead hand of statistics'* in the words of one respondent. The implications behind such discussions were that whereas the use of statistical methods allows no judgement, the simple procedures are not constrained by the protocols of statistical theory and so allow the auditor much more freedom.

The two firms who said that APs were a quality process considered that APs allowed the auditor to exploit their knowledge of the inter relationships within a set of financial statements and of the entity they represent.

6.3.7 Concerns about the use of APs.

As part of the interviews all the firms were asked if they had concerns about the use of APs. Three partners claimed that the use of statistically based APs was not appropriate since the accounting data were not stable over time. Procedures that require data from a period of time extending beyond two or so years were not used very frequently. Many of the Technical Partners expressed the view that trends were not stable over an extended

period (more than 2 years). When questioned on this, it was clear that they meant that statistical methods require data to be consistent over time, but they considered that in the environment of financial business systems this was not a feature that they possessed. They did not perceive that the same problems would exist for other forms of AP. Of course where the AP uses cross-sectional relationships between different data, they may be right. But any AP looking at longitudinal data has the same requirement for the relationships remaining stable.

Apart from that, all partners said they had no concerns, but issues were raised elsewhere in the interviews. Two firms were worried about the problem of obtaining adequate and reasoned explanations for the differences between the predicted and book values. Two firms feared that auditors did not understand business sufficiently to produce sensible and justifiable relationships. Such an understanding is essential; the Auditing Standards say that APs are about analysis of relationships between data both financial and non-financial data. They go on to say that it could be between different entities. All of this implies that there should be an understanding of the business.

6.3.8 Computer Tools.

One partner put forward the view that APs have improved over the last five years, and gave as the major reason the availability of powerful computers and software. The survey revealed that all the firms use computers to assist their APs. The software used ranges from spreadsheets through to statistical and bespoke audit software. The distinction between the big six and the second division firms is that all the big six firms use statistical software, but usually only for a few audits. For example, one firm uses a regression package they originally developed in the 1970's, and two firms use discriminant analysis. All the firms interviewed use spreadsheets as an analysis tool.

An interesting insight into the thinking on the subject of quality control in one of the firms interviewed in this survey arose when discussing the checking of spreadsheets. They saw no necessity to have any quality control measures on spreadsheets used to provide audit evidence. This is of concern since the use of spreadsheets, without appropriate checks can lead to false or misleading results.

That such checks are necessary can be illustrated by an example. Spreadsheets are used by organisations as a convenient and practical way of carrying out calculations; particularly when there are many different stages to a calculation. Consider the use of a spreadsheet to by an organisation, for which I worked, to compute a particular charge to a third party that was made up of different components. The person doing the spreadsheet had to bring together six sources of information, representing different parts of a service, to produce a consolidated charge. The results of the six separate calculations were supposed to be added to produce the consolidated charge. Unfortunately the summation was only done over five of the components. This led to the under-claiming of a large sum of money.

In the light of the potential problems illustrated in this example, the firm put quality control measures in place. This is an example of the advantages, to both parties, of face-to-face interviews.

6.4 Interviews with some Audit Firms in the USA

In 1994 as an adjunct to the survey of the 16 largest firms in the UK we visited the USA headquarters of four of the largest firms. A fifth firm was visited in 1998. The visits were partly on the recommendation of the UK firms. Another reason was to crosscheck policy; since for many of the big firms, the USA partnerships take the lead on policy issues. For one firm all the policy work is done in the USA, often with staff from the UK seconded to the teams. The same questionnaire was used in the UK and the USA.

Most of the results were similar to those reported above for the UK partnerships. There were two issues raised: documentation and the excessive use of simple APs by some firms. These are reported below.

6.4.1 Documentation

As in the UK, most firms claimed to document their APs in accordance with the Auditing Standards. Two of the US firms said that on legal advice they documented as little as possible. This was to minimize the information that could be obtained by the lawyers of those challenging their audit work. They recognised that this meant that their working papers did not meet Auditing Standards in the USA. The AICPA Auditing Standards have

extensive guidance on the format and content of the audit working papers. For example: *'Working Papers ordinarily should include documentation showing that: The audit evidence obtained, the auditing procedures applied, and the testing performed have provided sufficient competent evidential matter to afford a reasonable basis for an opinion, indicating observance of the third standard of field work'.* (SAS 41: Working Papers). The third standard of fieldwork is: *'Sufficient competent evidential matter is to be obtained through inspection, observation, inquiries, and confirmation to afford a reasonable basis for an opinion regarding the financial statements under audit'* (SAS 1).

The UK Auditing Standards have a similar requirement: *'Auditors should record in their working papers their reasoning on all significant matters which require the exercise of judgment, and their conclusions thereon'*. (SAS 230.3: Working papers). However, this is qualified in the next paragraph: *'The extent of working papers is a matter of professional judgment since it is neither necessary nor practical to document every matter auditors consider. Auditors base their judgment as to the extent of working papers upon what would be necessary to provide an experienced auditor, with no previous connection with the audit, with an understanding of the work performed and the basis of the decisions taken. However, even then, that experienced auditor may only be able to obtain a comprehensive understanding of all aspects of the audit by discussing them with the auditors who prepared them'* (SAS 230: Working papers).

The UK standards are less prescriptive, the last sentence quoted above providing the auditor with a reason to document very little. The UK firms all said they met the Auditing Standards, but given the complaints of the JMU described in Chapter 4, it is clear that the documentation is not good here. The difference is one of interpretation of what the Standards say. One UK firm said it was policy to remove most of the details about testing once the files were cleared by the senior partner in charge of the audit: they considered they were in compliance with the Standards. The result for audit documentation was the same as the USA partnerships.

6.4.2 A concern about the use of simple APs

For two of the firms, the use of simple APs by their competitors was a concern. The reason given, independently of each other, was the same. It was the fear that smaller firms were using inappropriate simple APs, thereby minimizing their costs and undercutting the fees of the large firms. Those large firms asserted that they maintained high standards of audit. These remarks arose in response to the question relating to their concerns about what other audit firms were doing. If we look at Figure 5.10 and Figure 5.17 we see that all the firms make great use of simple APs at the substantive stage, but that the larger firms make greater use of the more complex APs than the medium and small firms.

We found that the 16 largest firms do use the reasonableness tests and regression type procedures in the in the way that Fraser et al and Mulligan and Inkster found. But what is striking in our survey is that two of the big firms make much more extensive use of the 'more complex' APs than the others. Therefore one of the issues confronting those using these methods is that they feel disadvantaged because of the fee pressure. While it is to their credit that, in the mid 1990's, they still used the more expensive methods because they believed them to be of better quality, there are worrying implications for the fee pressure the firms are under and the consequent risk to the quality of work.

6.5 Conclusions

This survey represented an in-depth view of the important group of auditors who are responsible for developing and disseminating the audit methodology of the firms. Because of the interactive nature of face-to-face interviews this survey was able to explore issues, that a postal survey could not, and to gain a deeper understanding of questions underlying the application of APs. In particular, we were able to probe issues relating to the investigation rules (e.g. the data), the outcome rules (e.g. the explanation effect problem) and what evidence is gathered at the substantive phase of the audit cycle.

The survey reported in this chapter cannot be generally extrapolated to all auditors; it targeted a particular group: the Technical Partners of the 16 largest audit firms in the UK, or rather what were the 16 largest firms in 1994. Nevertheless it complements the other

British surveys: those by Fraser, Hatherley and Lin (1997) and by Mulligan and Inkster (1999). Taken together they present a picture of the application of a key audit procedure in the 1990s.

6.5.1 APs are an important audit procedure

Clearly APs are an important source of substantive audit evidence. All the surveys confirm that APs are used a great deal, that their use has increased and that they provide a significant proportion of the total substantive audit evidence. Simple APs such as simple trend analysis, ratio analysis and scanning predominate, and as we saw from Chapter 5, APs are claimed to provide the first signal of over 40% of all errors discovered by the auditor. They are considered to be a powerful tool, which may be a misconception, since APs appear to be used for about 40 percent of the total time charged to an audit. So APs are performing in proportion to the effort applied to them, and thus may not be quite so efficient as they are perceived to be!

6.5.2 APs are inexpensive

Auditors considered APs to be inexpensive to apply. There did not seem to be concerns amongst many of the audit firms about the quality of the data or of the rigour of the calculations. These two statements are consistent, since if more attention were paid to the data, APs might cease to be inexpensive procedures. The use of more disaggregated data might produce more precise results but would be more costly.

There is a possible contradiction between the reports that APs are inexpensive and that APs are used for between 30% and 40% of the total audit time. There are many interacting factors that have not been explored since they are not relevant to the main thrust of the work for this thesis. It could be, for example, that charge-out rates are behind this apparent contradiction. Partners may not get involved in the AP work, and relatively low cost staff may do that work.

6.5.3 Do auditors understand APs?

Auditors like to have flexibility to use their audit judgement; one of the Technical Partners interviewed in the survey said, *'The use of statistical type methods impaired and restricted my audit judgement'*. Although not all respondents expressed their views as frankly as this, it does point to a perception that audit procedures should not be fettered by 'rules' that auditors consider restrict their ability to apply judgement. The use of the decision-trees, and scoring systems, to assist in the decision about what type of substantive evidence to gather might appear to contradict this. However such aids are used to bring together the auditors' judgements. For example whether a particular audit procedure when considered in conjunction with other audit work, such as work on the systems of the audited organisation, is adequate to have a reasonable chance of detecting material error. The decision-tree, or scoring system, advises what type of substantive evidence to collect!

Although many of the respondents considered the nature of the substantive evidence, there does not always appear to be consideration of the extent of the substantive evidence that is required. This phenomenon can be illustrated by looking at one firm's approach. Although it concerns sampling, similar considerations apply to APs. The firm sets materiality for an audit but does not then link the amount of audit evidence to that value. For example, they calculated that a sample size of 300 was required when sampling to achieve a materiality level set at one percent of a particular value. That was judged to be too much work and so the firm's audit policy allowed the auditor to restrict the sample to 30 'items'. In the partner's opinion, there was nothing to be gained from looking at more cases than this: *"If something is wrong, then a sample of 30 would reveal that"*.

This example was given by a respondent in the survey to illustrate the policy of their firm and is repeated here because it is a good illustration of one of the issues surrounding auditors and evidence: the link between the amount of work they do and the outcome of that work is not always explicit. It was a surprising admission since there is a link between sample size and the information gained. Such a link should be obvious to the auditor. The Auditing Standards recognise it in the SAS on sampling, which says: *'Tolerable error is considered during the planning stage and, for substantive procedures, is related to the*

auditors' judgment about materiality. The smaller the tolerable error, the larger is the sample size as a proportion of the population' (SAS 430). The link can be expressed in the following way: for a given sampling plan, the larger the sample the less uncertainty and/or imprecision associated with the results of that sample. Some sampling schemes may be more appropriate for a given set of circumstances than others: such choices are judgements made to meet the objectives of the particular audit test and the efficiency criteria of the auditor.

For APs similar considerations apply. The type of AP (investigation-rule) chosen and the quantity and quality of the data will have implications for the accuracy of the outcome of the AP. The better the investigation-rule, in the sense that it more accurately models (reflects) the true position, the less false signals there should be. The outcome-rule (outcome-value) also has an impact. The smaller the value, the surer the auditor can be that the value being audited is correct. To illustrate this, consider an absurd case when the outcome-value is close to the figure being audited. Then almost any investigation-rule will succeed in the sense that it will confirm to the auditor that there is no significant error: albeit that the result would be ambiguous. Thus for APs there is a link between the amount of audit work: selecting the data, checking of that data and building of the investigation-rule and the consequent quality of the outcome.

One observation made by Humphrey and Mozier (1990) was that auditors appeared not to understand the procedures they used. After reviewing the surveys in Chapter Six and considering the survey described in this Chapter, what could we say now about that understanding?

An understanding of the process would be demonstrated if there was a structure to the AP process. Using the vocabulary (terminology) of this thesis, the structure would include an investigation-rule, an outcome-rule (together with an outcome-value) and a procedure for obtaining explanations. Such a set of "rules" would not, necessarily, require the so-called "advanced" statistical methods such as regression. But would require a logical structure (perhaps in sentence form) to the process. Such logic would cover simple comparisons as well as more complex investigation-rules. The structure would not fetter, or impair,

judgement, but would make clear those judgements. A requirement to set down the investigation-rule, outcome-rule and to consider and document alternative explanations for differences between the prediction(s) and the value(s) being tested would make the judgements transparent and guard against accusations that there was no independence to the process. Such a structure should not make the process bureaucratic since it is only making clear what auditors claim they already do when they use their knowledge and audit-judgement.

The views that statistical methods or theory get in the way of judgement are not founded in any logic. It is as through the auditor does not want an objective result, but one that they can manipulate. It is likely that the manipulation is not one of corruption, but more that they do not see the need to be in a position to explain or justify their results to a third party. The distrust of theory is epitomized by the quote about the Audit Risk Model given in Chapter 4 from York (1997): '*the sooner it is returned to the box labelled "theory - dangerous if taken too seriously" the better*'. Such sentiments were typical of several of those interviewed in the survey reported in this chapter, e.g. the auditor's lack of consideration of data quality and the failure to validate the explanations they receive from management.

In section 6.3.2 we saw that many auditors fail to define an outcome-value for the outcome-rule $| \hat{Y}_T - Y_B | \leq C$. They said that if the expectation value $\hat{Y}_T$, was close to the book value in the financial statements being audited, then that would be 'the right answer', but did not define close in advance. One interpretation of this is that APs have a confirmatory role but not an exploratory role.

Another concern of two firms was that auditors did not sufficiently understand the business. The need to appreciate the inter-relations between data from different parts of the entity or similar entities underlies APs. As the Auditing Standard on APs says, one of the factors in the application of APs is: *'the knowledge gained during previous audits, together with the auditors' understanding of the effectiveness of the accounting and internal control systems and the types of problems that in prior periods have given rise to accounting adjustments'* (SAS 410).

We saw in section 6.3.3, that there was little or no consideration of explanation effect phenomenon and the willingness, observed in Chapter 4, for auditors to accept management representations too easily - Boys (1997).

An event outside of the research for this thesis into the surveys of auditors' views illustrates the thinking of some in the audit profession. In 1999 the Scottish Institute of Chartered Accountants submitted a revised syllabus to the DTI. They wishes to drop the subject of "quantified methods" considering that such methods had no relevance to auditors. One of my supervisors, who worked as a consultant to the DTI sought my views on this. We agreed that since APs are used extensively, auditor should be aware of the thought processes associated with quantified methods even if they are not choosing to use methods such as regression. They should make that choice based on knowledge.

For many of the auditors in the survey, a structure to the AP process, or at least parts of it, does not exist. So Humphrey and Mozier's concern appears to still have some resonance today. That is auditors do not understand the process underlying a procedure such as APs.

6.5.4 An area for further exploration: nugatory work

In Chapter 3, we raised the problem that false-negative signals could cause the auditor to undertake nugatory work. It would assist our understanding of the effectiveness of APs to have reliable evidence from the audit firms about whether all the signals of potential error are proved to be correct. That is, how many signals of error are there that subsequent investigation reveals are false signals? It is unfortunate that none of the surveys have explored this particular issue. The survey carried out for this thesis, and described in this chapter, did touch on this issue, but it was not feasible to carry out a detailed investigation. While the inspection of one audit where this occurred would provide some insight, what is needed is the information from a number of different audits: by different auditors and firms. Such evidence would be difficult to collect and verify; it would require the co-operation of several audit firms and would involve asking auditors to record the information on a large number of audits that they undertook. Given the criticism of auditor's documentation by Garvey and Dietz (1997) and others, as recorded in Chapter 4, much effort would be necessary to persuade auditors to open up commercial information.

6.5.5 An area for further exploration: can a simple AP be effective?

There is a hypothesis that can be built on the information gathered in this survey. It flows from the extensive use of simple APs. One might conclude that 'APs are used a lot to collect small, but important, amounts of evidence at a small cost to the auditor'. The philosophy of the audit approach illustrated by the tree-diagram discussed in section 6.3.5 and in Figure 6.6, which drives the auditor down the path of identifying risks, looking for controls to mitigate those risks and then looking for a procedure that is cost effective (i.e. inexpensive to administer) and meets the requirement to do some substantive work. Simple APs meet the requirement for the auditor to be seen to be complying with the SAS. *'The assessed levels of inherent and control risks cannot be sufficiently low to eliminate the need for auditors to perform any substantive procedures for material account balances and transaction classes. However, these substantive procedures may comprise only analytical procedures where such procedures provide sufficient appropriate evidence'* (SAS 300).

We are not arguing against the strategy of identifying the riskier parts of the financial statements being audited, and addressing those risks. This thesis addresses the question of whether simple APs are capable of providing evidence to confirm or disprove the existence of error. Where there is low risk (perceived) it may be very reasonable to apply the minimum level of substantive testing. This may be a simple AP. If that confirms the low risk (i.e. no material error identified) that achieves the auditors' aim provided it is powerful enough to contradict the evidence from the auditor's controls assessment when that is appropriate.

6.5.6 Key results

The key results from this chapter are that APs are confirmed to be an important substantive procedure and that simple APs are the most commonly used APs. This chapter has additionally identified some concerns about the understanding of those procedures. These concerns include a lack of attention issues such as data quality, a lack of an active consideration of the explanation-effect problems and the fact that the use of explicit

outcome-rules is not always considered. There is an apparent lack of structure to the AP process.

Given this concern, the rest of the thesis addresses the issue of whether APs can detect material error in a consistent way, without excessive levels of false signals. Chapter Seven will review past experiments into the effectiveness of APs and Chapter 8 will extend that work with some new experiments that investigate the frequencies of false signals.

Chapter 7: Investigations of whether APs correctly signal the presence or absence of error

7.1 Introduction

A vital attribute for an AP is its ability to signal correctly the presence or absence of material error in a set of financial statements. This chapter reports on the work of some researchers who investigated whether particular APs correctly and effectively signal error. That APs do so is important since they frequently replace other forms of substantive tests, such as sampling, and thus may provide the only substantive evidence for a particular assertion. In this chapter we examine the experiments of others that investigated whether particular APs correctly signalled the presence or absence of material error.

Chapters 5 and 6 reported surveys that investigated the use of APs by auditors. As a consequence we know that APs are seen as an important procedure at the substantive stage of an audit. The result from the application of a substantive AP is one of the signals that the auditor uses to conclude whether there is significant (material) error in the part of the financial statements being audited. For the purpose of this thesis, the important question is how can we be sure that an AP is providing a reliable signal? That is, if material error exists in the audited population, then the AP signals accordingly with a high probability. Similarly, if there is no error the AP should correctly signal this with a high probability. So we require low frequencies of false-negative and false-positive signals.

7.2 Outcome-rules: What have other researchers used?

In Chapter 3 we defined APs by two rules: an investigation-rule and an outcome-rule. Once the application of the investigation rule has produced the expectation-value, $\hat{Y}_T$, we need some criteria to measure how close it is to the value being audited Y_B (the book value). That is, we need a rule for assessing the closeness of $\hat{Y}_T$ and Y_B. We defined this, earlier, as the outcome-rule $|\hat{Y}_T - Y_B| \leq C$, where C is some specified outcome-value.

A wide variety of rules have been used. Coakley (1982) identified three categories of outcome-rule used to determine whether the deviation of the expectation value from the

actual balance is significant. They are a useful categorisation when discussing the different approaches and can be summarised in terms of the outcome-value (C). For each of the categories below the test is: $|\hat{Y}_T - Y_B| \leq C$.

1. For judgement based rules C might be an arbitrary threshold. It would be considered arbitrary in the sense that it is not formally linked to any other value such as materiality. It may well be based on the auditor's judgement. For example, when comparing balances the threshold might be 10% of the previous balance. In the Passport example a value of £9m (10%) might be applied to the 'Income from the issue of passports' line (value £90.5m).

2. For materiality-based rules, C is the overall materiality or some function of it. In the Passport example a value of £1.075m might be applied to the Income from the 'Issue of Passports' line.

3. For statistically based rules, C is a specified multiple of the standard deviation associated with the prediction. An example of a statistical rule might be a regression based prediction interval based on a time-series associated with $\prod_T$, for which there is a probability $(1 - \alpha_o)$ that a future single observation on y_i, given $x = x_i$ will lie within the interval: $\hat{Y}_T \pm t_{\alpha/2,\, n-2}\, s \{1 + 1/n + (x_i - \text{xbar})^2/\sum(x_j - \text{xbar})^2 \}^{1/2}$. Details of the symbols are in section 3.8.2.

Although the Coakley rules are an over simplification, they are still useful. Rule one does not explicitly use the auditor's materiality. If we consider the requirements in the Standards for the auditor to have regard to the overall materiality when doing each test, then rules one and two should be the same. There should be a rationale associated with rule one that does have some regard to 'materiality'. Where auditors appear attracted to rule one, with its apparent disregard for materiality, it may reflect their desire: *'to use their professional judgement'* as reported in Chapter 6. Rule three, the statistically based rule, as defined, takes no regard of the auditor's materiality value, although it could easily be adapted to do so.

All three categories of outcome-rule have been used in the experiments surveyed (covering the period since 1975). For example Loebbecke and Steinbart's (1987) research into the effectiveness of very simple APs used the judgement type rule: it was

to compare the current year balance to that of last year. The rule triggered an investigation if the comparison noted a change of more than 10%. Loebbecke and Steinbart found that outcome-rules of 10% to 15% failed to signal the presence of error. As seen in Chapters 5 and 6, the simple APs used by Loebbecke and Steinbart are the most commonly applied APs.

An example of the application of Coakley's rule two is Knechel's investigation of non-statistical tests with materiality-based rules. He used rules such as: *'investigate any month which exceeds by a material amount the equivalent prior month allowing for anticipated changes in the market or company performance'* Knechel (1988b). Suppose we are comparing the performance of each of the months from this year with last year. In the audit year we have the values: $Y_{2,1}$, $Y_{2,2}$, $Y_{2,3}$, .., $Y_{2,8}$, ..$Y_{2,12}$. In the prior year we have the values: $Y_{1,1}$, $Y_{1,2}$, $Y_{1,3}$, …, $Y_{1,9}$, ..$Y_{1,12}$. One such comparison would involve month three from each year: $Y_{2,3}$ versus $Y_{1,3}$. Then the outcome-rule might be $|Y_{2,3} - Y_{1,3}| \leq C$, where C, the outcome-value, is Knechel's material amount.

Several pieces of research looked at the effectiveness of APs using the statistical rules in conjunction with materiality. These included those of Stringer (1975), Kinney (1979), Kinney, Salamon, and Uecker (1986) and Knechel (1988a). They all studied expectations based on regression with statistical outcome-rules.

7.3 Methods for investigating whether APs are effective

Once we have set up the investigation-rule and the outcome-rule, we need to test whether the AP will detect error. One way to obtain the evidence that APs work would be to inspect actual audit files. For these to be useful we would need to know everything about those populations, including the book values and the errors. In effect, we would require a 100% audit of those financial statements. Given the size of the populations that underlie financial statements, it would be impractical to obtain such information. Even if we use high-level data, a problem would be to collect a large number of such populations. We therefore need to adopt another strategy, based on simulation.

7.3.1 Simulation

The Concise Oxford Dictionary, (tenth edition - 1999) defines simulate to be: '*imitate or reproduce the appearance, characteristics or conditions of*'. In its technical sense simulation involves using a model to produce results, rather than experimenting with the real system under study. If the model has a stochastic element, it is sometimes known as stochastic simulation. The Monte-Carlo method arose during Second World War for stochastic simulations of models of atomic collisions; the term is frequently used synonymously with stochastic simulation.

For statisticians, the term simulation describes a variety of techniques, which mimic the rules of a model. When the model can be given a mathematical formulation, but analytic predictions are not possible, then simulation may prove a useful tool to investigate how the model behaves under changing circumstances. There have been many applications to real life situations. For example, following the 1979 crash of a DC-10 Jet after it lost an engine, the Observer newspaper (10 June 1979) reported that, *'A DC-10 flight training simulator is being programmed to investigate whether the aircraft would be controllable with one engine detached'.*

The Millennium Bridge in London was designed with the aid of computer simulation. It was not stable and swayed when too many people crossed it: a possible cause of this failure was that the computer model was incomplete. This illustrates that the success of such simulations in predicting how an entity or system, the bridge in this case, will function depends on the model used and how well it describes the situation. Another example involves the scientists from Imperial College, who used simulation to investigate the effectiveness of different eradication policies by modelling the spread of the foot and mouth outbreak in 2001.

As it is not normally possible to obtain large sets of real data that represent the full population (unless we can take a census), we may need to use simulation techniques to create artificial populations, with similar characteristics to the full population. For this artificial population we would have complete information about values and error. Such data sets, whose characteristics are known, provide a basis for experiments to test whether a particular AP will detect error in a consistent way.

To model the application of an AP using simulation there are a number of key elements to the process. First we create sets of data where all the characteristics of the audit population are understood so that we can measure the success or otherwise of the AP. Then the AP is described in terms of the investigation rule and the outcome-rule. This can be done through case studies or computer simulation.

7.3.2 Case Study Approach

One way to prove that APs do or do not reliably signal error would be to construct case studies and use large numbers of 'auditors' to carry out the procedures under controlled conditions. There are a number of steps involved in such a case study approach. First we need to collect suitable volunteers to play the role of 'auditors'. Ideally they would be from the population of practising auditors in order to replicate the audit decision process. Then data would be collected or simulated. The next step would be to construct a case study scenario for the AP. As part of this scenario the audit would be described so that the auditors had all the relevant information for them to devise and implement the AP. Included in this information would be the outcome-values to be set for measuring the performance of the simulated AP. The question could be: does the deviation between the expectation value and the figure in the financial statements exceed the outcome-value? For example, using judgement-based rules the outcome-value could be some threshold; whereas for materiality based rules the outcome-value could be related to the overall materiality or some local materiality. The 'auditors' then perform the APs and then the results are analysed.

A major problem with this type of approach is constructing a case study that allows for the variation in skills and experience of the people participating. If we are testing how people react to the data, and to the results, then the case study approach is an appropriate one. However, the objective in this thesis is to test whether different combinations of investigation-rules and outcome-rules are consistently signalling whether error is present. To achieve this objective the distributions of the results need to be modelled. For this a large number of replications must be performed, making the case study approach impractical. We therefore turn to computer simulation, as described in the next section.

7.3.3 Computer Simulation

The use of a computer as an aid to simulation enables many thousands of trials to be done in an economical way, providing the structure of the simulation is efficient. In computer simulation the usual structure is to formulate the problem, define the model and collect the data. A computer program is then written, and the programming checked. The next step is to carry out pilot runs and verify that the model and the programs represent the problem that is being solved. Once that verification is done the programs are run and the results analysed to produce a solution to the problem. In reality this formulation is the same as solving any problem, but the quality checking of the model and the associated programming is essential to ensure that it addresses the problem posed. *'If a model is not a 'valid' representation of a system under study, the simulation results, no matter how impressive they appear, will provide little information about the actual system'* (Law and Kelton (1991)). Subject to this caution, Law and Kelton point out several advantages including one that is particularly relevant to the experiments in this thesis: *'In simulation we can maintain much better control over experimental conditions than would generally be possible when experimenting with the system itself'.*

7.4 An introduction to some investigations into the effectiveness of APs

Some of the research published recently has described simulation experiments into the performance of a number of more complex APs; examples include multivariate procedures such as the ARIMA and X11 models. Examples include Ameen (1989), Dzeng (1991) and Chen and Leitch (1998).

Ameen (1989) compared an X-11 model against a regression-based model and found the former to be superior. The X-11 model is a time series model that decomposes time series data into trend-cycle, seasonal and irregular components. Ameen found that for the data she used, the X-11 time series model had a significantly lower detection risk than a regression model and she considered it was much easier to use. The level of detection risk was assessed through the Audit Risk Model. Dzeng (1991) favoured a Bayesian Autoregression model, known as Vector Autoregression. Dzeng claimed it was easier to use than Multiple Regression because of the extensive testing of the base data that is required to satisfy the theoretical requirements of Multiple Regression. This

is a rather unfortunate conclusion since the application of any model should involve the consideration of the data and the appropriateness of the model to that set of data. Autoregression models require long time periods, typically 100 observations, which is rather long according to the opinion of the Audit Partners interviewed in the survey described in Chapter 6. Four years worth of data is longer than some Audit Partners thought would be stable. So the data requirements for Autoregressive methods may not be met, although to some extent this comment applies to other APs as well.

Such models are not typical of the types of AP that the surveys have shown that auditors use. Therefore we will concentrate on the range of procedures that auditors use extensively to ascertain the effectiveness of those procedures. In the remainder of this chapter we therefore describe four pieces of research focusing on simple APs (in sections 7.5 and 7.8), ratio analysis (in section 7.6) and regression (in section 7.7).

7.5 Do simple APs work?

There have been few published research papers that have focused on simple APs in the context of substantive audit. Of particular interest is the work of Loebbecke and Steinbart (1987), who focused on a set of simple planning type APs that they claimed were used by auditors to obtain substantive evidence. The methods included scanning, simple trend analysis and simple ratio analysis using annual, not disaggregated, data and utilised simple outcome-rules. The surveys described in Chapters 5 and 6 confirm Loebbecke and Steinbart's argument that such simple APs are extensively used to obtain substantive evidence.

Loebbecke and Steinbart's investigations related to financial statement data collected from publicly quoted companies. There were two samples of 38 and 22 companies, which were selected to ensure that they represented a wide variety of industries. Part of Loebbecke and Steinbart's rationale was that such simple APs *'appear to be applied to all audit clients'*. They do not describe their simulation process in detail, but it is clear that data from 1980 to 1983 was treated as prior years audited data, and data from 1984 was treated as the current years data that was to be audited. They stated, *'in performing the simulation we assumed that the 1984 data were correct. Therefore we defined a decision to investigate the 1984 reported amounts as a type I*

error' (false-negative signal). *'We then seeded a material error into the statements and defined a decision not to investigate as a type II error'* (false-positive signal).

The definition of 'material' used by Loebbecke and Steinbart was $1.6 \times X^{2/3}$, where X is the greater of total assets or revenue. Although such a rule is not scale invariant, it was consistently applied to the dollar values and so is not an issue. The objective of the experiments was to measure the effectiveness of the procedures in identifying error. They did this using the frequency of the false-negative and false-positive signals as a measure of the effectiveness.

The first experiment by Loebbecke and Steinbart used an investigation rule that compared the current year balance to that of the previous year and an outcome-rule such that if the change was more than 10%, the AP signalled that there was a potential material error. This is similar to 10% variation rule (outcome-rule) described by Coakley (1982). For the outcome-rule with an outcome-value of 10%, the range of false-negative signals was between 45% and 63% depending on the account area. The false-positive signals range between 24% and 97%. The details are shown in Table 7.1 below. The outcome-value was later changed to 5% and then 15% with similar results.

Account Area being tested	False-negative	False-positive
Account receivable	50%	45%
Allowance for doubtful accounts	63%	24%
Inventory	63%	42%
Accounts payable	45%	87%
Sales	55%	45%
Cost of Goods Sold	58%	97%
Selling, General and Admin Expenses	58%	97%

Table 7.1: The results of an experiment by Loebbecke and Steinbart (1987) showing a simple comparison type AP with an Outcome Rule = 10%. [Source: Computed from data within the Loebbecke and Steinbart article].

All of the outcome-rules exhibited a high rate of false-negative and false-positive signals, varying with the account area being tested: the false-negative signals being less dispersed across the different account areas than the false-positive signals. For the false-positive signals we see that when the AP was applied to the 'allowance for doubtful accounts' the performance was better than for other areas.

In the sense of having a lower frequency of false-positive signals, the APs did best when the material error was high relative to the account area being tested and therefore the error seeded was large relative to the account area being tested. Although the relative size of the error is only one factor in the performance of an AP, this result implies that unless the error is extremely large, errors that are material to the financial statements may be missed.

In a second analysis, Loebbecke and Steinbart used eleven different methods of generating an expected account balance, which were compared in terms of their predictive accuracy. One example of an investigation rule is: the expectation-value for year [i +1] is the value for year [i] plus the difference between the values in years [i] and [i – 1]. Another example is the use of regression, although the latter was on only four years annual balances! The results of the second analysis are shown in Table 7.2.

Account Area being tested	False-negative	False-positive
Account receivable	37%	74%
Allowance for doubtful accounts	50%	16%
Inventory	58%	50%
Accounts payable	55%	63%
Sales	16%	95%
Cost of Goods Sold	21%	92%
Selling, General and Admin Expenses	16%	95%

Table 7.2: This shows the best of 11 Predictors with an Outcome Rule = 10%. It also shows that for most account areas, there are high occurrences of false signals: both positive and negative. [**Source:** Computed from data from Loebbecke and Steinbart].

The decision rule was based on a measure of prediction (expectation value) error calculated by: (expectation value – actual value)/(expectation value). Loebbecke and Steinbart's (1987) reason for choosing the 'expectation value' as the denominator was explained as follows '*when conducting APs the auditor compares the unaudited actual values to their own expectation. The auditor should form the expectation before looking at the unaudited value, to avoid biasing their expectations by anchoring on the actual value'*. In Loebbecke and Steinbart's opinion, *'the important measure is how far the book value deviates from the auditor's expectation'*.

Loebbecke and Steinbart statement that the auditor should form the expectation before looking at the unaudited value is correct and is entirely consistent with the argument a the end of Chapter 6. There we argued that auditors in the UK do not do this, and this is a serious deficiency in the application of APs.

We can see that although these APs performed better than the very simple AP in Table 7.1, there were still high frequencies of false-negative and false-positive signals. Other than for the account area 'allowance for doubtful accounts' the levels of false-positive signals are very high. This all the more alarming if we recall that these results (Table 7.2) record the best performing of the different APs. We can see from these results that whether it should be the expectation-value [$\hat{Y}_T$] or the true value [Y_T] is, in this experiment, immaterial since the overall conclusion from Loebbecke and Steinbart's work would be the same. That is these APs were not good at detecting error.

For each account area in Table 7.3, the results are the optimal performance of the APs that according to Loebbecke and Steinbart '*minimized the average absolute error of prediction for an account, across all 38 firms*'. Also shown in Table 7.3 is data on the average change in each account area between the prior and current years. The prediction errors associated with the account areas varied between 5.4% and 16.5%. There was a large variation around these values. For example Loebbecke and Steinbart report that the standard deviations associated with these prediction errors values were respectively 5.3% and 11.6%.

Account Area being tested	Materiality as a % of the account area	Average absolute error of prediction	Average absolute % change from prior year
Account receivable	1.2%	10.3%	12.7%
Allowance for doubtful accounts	45.4%	14.9%	16.8%
Inventory	1.9%	16.5%	18.7%
Accounts payable	2.1%	15.0%	15.6%
Sales	0.1%	5.6%	10.8%
Cost of Goods Sold	0.2%	6.6%	11.3%
Selling, General and Admin Expenses	0.7%	5.4%	10.8%

Table 7.3: This shows that the average error of prediction associated with each account area was large. The standard deviations associated with these were large. The table also shows that the data used for the APs was not stable over time, exhibiting wide variations between years. [Source: Computed from data from Loebbecke and Steinbart].

Which of the eleven types of AP performed best is shown in Table 7.4. Clearly no one AP is the right one for all account areas. Even the 'optimal' APs are not performing well as can be seen if we crosscheck with Table 7.3.

	Method described by Loebbecke and Steinbart	The method performed optimally for account area:	Average Prediction Error
1	Last year's account balance		
2	Expectation-value for year [i +1] is the value for year [i] plus the difference between values in years [i] and [i – 1],	Sales Cost of goods sold	5.6% 6.6%
3	Simple time series regression on last 4 years	Selling, General and Admin Expenses	5.4%
4	Model relating account areas		
5	Average change over last 4 years added to last year's balance	Allowance for Doubtful accounts. Accounts payable	14.9% 15.0%
6	As 5, but use weighted average	Accounts Receivable	10.3%
7	Simple average of previous 4 years		
8	Weighted average of previous 4 years		
9	Similar to 2, but use percentage change		
10	Compute the average percentage change in account balance	Inventory	16.5%
11	Repeat 10, but use a weighted average percentage change		

Table 7.4: No one AP will perform best for all accounts areas. [Source: Loebbecke and Steinbart (1987)]

These experiments indicate that, potentially, such APs are poor indicators of material error. Loebbecke and Steinbart (1987) concluded that the simple APs tested did not reliably detect material error: '*As can be seen, for almost every account both the error of prediction and the average absolute percentage change in the account swamp the effect of adding a material amount or subtracting a material amount from the account balance*'. In other words, as well as the large error of prediction found, the volatility of the account areas would mean that material error could easily go undetected. Indeed this implies that many account areas are not stable enough for APs to be applied to them. If this finding were repeated, it would invalidate the application of this type of simple AP. The results corroborate an opinion expressed by some of the Audit Partners in the survey reported in Chapter 6, that data may not be stable over a period of more than two years. The SAS says that: *'the application of analytical procedures is based on the expectation that relationships between data exist and continue in the absence of known conditions to the contrary'*. This implies that over the short term, the relationships between data remain stable. So if such volatility as was observed with the

data used by Loebbecke and Steinbart were repeated, then this requirement from the Auditing Standards would not be achieved. The volatility may have been associated with the higher inflation that existed in the early 1980s.

As a consequence of their experiments Loebbecke and Steinbart recommended that such procedures should not be used to provide substantive evidence, either as a means of reducing or replacing other audit tests.

7.6 The use of Ratios as a substantive AP

One of the simple types of APs tested by Loebbecke and Steinbart (1987) was ratio analysis: it did not do well. Whittington (1980), Barnes (1982) and Lee (1985) and others have argued that, in a financial setting, there is evidence that the relationship between numerator and denominator variables in the simple ratio format is not strictly proportional. If the form of the relationships for the two variables is $y_i = \alpha + \beta x_i + \varepsilon_i$, they are arguing that $\alpha \neq 0$. For proportionality, $\alpha = 0$. This is important when considering the ratio form, which is obtained by dividing through by x_i to give: $y_i/x_i = \alpha/x_i + \beta + \varepsilon_i/x_i$. If $\alpha \neq 0$, the ratio is skewed by α/x_i, and if α is positive, the ratio will decrease as x increases.

A frequent use of ratios is to measure the ratio against a norm. That involves the simple calculation of, and comparison against, industry averages. Such comparisons may be based on an assumption of normality. The auditor needs to be aware of this and consider the median and mode as well as the mean (average). The simple measures of dispersion, such as standard deviations about the mean may lead to erroneous conclusions.

The result of two studies by McDonald and Morris (1984, 1985) concluded that the ratio form is appropriate for financial statement analysis. Berry and Nix (1991) argued that this conclusion was 'too strongly stated'. They posed three questions, which revolved around whether the findings of McDonald and Morris [i] would generalise to other industries, [ii] would be stable over time and [iii] would hold for other ratios. Without positive answers to these three questions, Berry and Nix (1991) asserted that: *'analysts would do better to adopt an approach to financial statement analysis which can identify the presence of, and cope with, a wider range of circumstances than those adequately dealt with by the ratio form'*.

Berry and Nix addressed these questions, using regression to model the relationship between each pair of variables y_i and x_i that would typically be linked by a ratio. First they applied a regression model of the form $y_i = \alpha + \beta x_i + \varepsilon_i$, and then after dividing through by x_i, applied regression on $y_i/x_i = \alpha/x_i + \beta + \varepsilon_i/x_i$. They used as the data for the experiments financial information from the brewing industry, extracting information on 37 brewing companies from a corporate financial database: Microexstat covering fifty different industrial classifications. They obtained the financial information for four consecutive years. From this they extracted the relevant data for five ratios: current assets to sales, current assets to current liabilities, cash flow to total debt, total debt to total assets and stock to sales. The first four ratios were those used by McDonald and Morris (1984, 1985); the fifth was an extra ratio added by Berry and Nix.

As a result of the experiments, Berry and Nix found that: *'This study offers little support for the finding of McDonald and Morris'*. In particular, analysis of the data from the brewing industry indicated that McDonald and Morris's finding, which was that the ratio form is appropriate for financial statement analysis, did not transfer to other industry groups. The information on which to base this statement is limited: McDonald and Morris's and Berry and Nix's experiments each only had one industry group, although Berry and Nix seem to have found a counter-example to the findings of McDonald and Morris at the first try.

Because the basic assumption of proportionality (i.e. $\alpha = 0$) of simple ratio analysis is usually violated, an alternative considered by Berry and Nix was the use of regression models. These they argued are more appropriate than ratio analysis, but more complex. Others, such as Whittington (1980) do not consider regression methods to be beyond the auditor in this age of laptop computers with relatively simple software packages to assist the auditor. The findings of Stewart (1978) certainly support that argument.

Berry and Nix, who were looking for some other procedure, concluded that: *'It does appear that there are variables commonly combined as ratios where that is inappropriate'*, and *'For the variable pairing of stock and sales, different models proved best in different years. An assumption free method of analysis is therefore desirable'*. To address this problem, Berry and Nix advocated something to replace regression analysis that is as simple to use as ratio analysis. Although Berry and Nix

did not specify what that 'new' technique should be, they proposed that an assumption free method of analysis would be desirable. One approach they suggested was the exploratory use of regression. In an article on Initial Analysis of Data (IDA), Chatfield (1985) provides an excellent discussion of the use of regression in this way.

Ezzamel and Mar-Molinero (1991) proposed a similar approach; their method relied on the Exploratory Data Analysis (EDA) methods developed by Tukey (1977). Both Chatfield's and Tukey's methods are similar and they are concerned with understanding the data as the first step in any analysis, possibly using methods without too much concern for theoretical assumptions about those methods. A difference between IDA and EDA is that the latter is more formal. An approach such as IDA or EDA would not be appropriate for the formal analysis. The IDA concept was developed because of the perceived dangers of using statistical computer based analysis packages: that without a proper understanding of the data, an inappropriate method might be applied to the data or there could be errors in the data that would not be identified because the analyst did not inspect the data. Understanding the data is a key aspect of addressing the problem. However it must always be borne in mind that no technique is likely to work if the relationships are unstable. Statistical methods only work if there is replication, which implies that there is some form of stable relationship.

A simple example of the problems that can arise to which Tukey and Chatfield were drawing attention arose in the course of the author's investigation of traffic accidents in the County of Avon. The police could not understand why the average age of a male involved in an accident was 22 years and of a female was 75 years. On examining the data, the explanation for this phenomenon became obvious. The default code for 'will not give their age' was 99. More women than men had refused to give their ages. The police just fed the data into the computer, and analysed it with a statistical package, which gave the appropriate answer for the data it was given. But it gave an incorrect answer to the question: what are the average ages of male and female drivers involved in an accident? Because of the large apparent discrepancy between the ages of males and females, the analyst queried the answer. If fewer of the ladies, or more of the men, had not given their ages then the discrepancy between the average age of male and female drivers would have been less obvious and the solution provided by the package may not have been queried. Although in one sense this is a trivial example, the

principle applies to all analysis, including APs. In another analysis the consequences could be important, for example, in an auditor's attestations. Although the concepts of IDA were directed at statisticians, they are equally relevant for any analyst.

If regression is not acceptable, being seen as beyond straightforward application by auditors, another solution identified by Berry and Nix might be to use ratio analysis, but with the comparison not to a single target value, but to a range of appropriate target values. The example they give is that the ratio of stock to sales for a given firm would not be compared to an industry average but to an average of small or medium or large firms in the industry. The survey described in Chapter 6 asked the audit firms about this. None of the firms claimed to use such procedures, although several made the point that auditors intuitively do this when considering the financial statements. It may be that auditors do informal modelling founded on their experience and knowledge of the client.

A better solution might be to compare not one ratio in isolation, but all of the relevant ratios to the appropriate industry averages. If this was true for all $|\hat{Y}_{Ti} - Y_{Bi}| \leq C_i$, then this might be strong evidence. But the difficulty in evaluation would be if some were > C_i and some were not.

However the survey also found that over one-third of the firms did not check the data, many did not connect the overall materiality to the outcome-value in any way and one-third had no procedures in place to ensure that explanations were obtained if $|\hat{Y}_T - Y_B| > C$, where C is the outcome-value. That provides little encouragement that auditors have adopted, or will adopt, the methods that Berry and Nix perceived that some analysts were using: comparison with several variables. Berry and Nix stated, *'This approach is current in at least one major UK bank'*. But if some auditors do not, or cannot, cope with one explanation it is hard to see how they would deal with several. In comparing to several values, the problem would be measuring acceptability of the expectation value against that criterion. For example, is the vector of the different comparators (C_i), in the intervals $|\hat{Y}_{Ti} - Y_{Bi}|$? Auditors might be able intuitively to assess all these variables, but they need to demonstrate that they can repeat that assessment, and document it for others to replicate.

To sum up, Berry and Nix found evidence that the three questions set out on page 171 were not answered in the affirmative. Firstly, the findings of McDonald and Morris (1984, 1985) do not generalise to all other industries, a counter example found by Berry and Nix being the brewing industry. Secondly, ratios are not stable over time: *'For the variable pairing of stock and sales, different models proved best in different years'* (Berry and Nix 1991). The auditor would not have the luxury of knowing which models are performing best, since they do not know the true value. Thirdly the ratio that Berry and Nix used was additional to those of McDonald and Morris. It was stock to sales and it did not perform well. Thus this study supports the conclusion of Loebbecke and Steinbart (1987), which is: simple APs (e.g. ratios) are not appropriate for substantive testing. Although this is true for cross-sectional data such as those used by Berry and Nix (1991), separate analysis time-series data is desirable. A study by Knechel (1988b) will consider this. The description is deferred to section 7.8 because the evaluation procedure used by Knechel was similar to others to be described in section 7.7.

Berry and Nix identified a need for some non-parametric procedure. Neither IDA nor EDA is likely to prove an acceptable procedure if the requirement is to produce substantive audit evidence, although they could play a role as audit planning APs. Another procedure not considered by Berry and Nix is non-parametric regression. As Silverman (1985) commented when advocating such methods: *'there may be some advantage in letting the data determine the fitted model in somewhat more flexible way'* than Multiple Regression or Time Series models. Such models can be applied to explain the current year's data, to provide a basis for comparisons between years and to build a model of the process. They require powerful computers and these are readily available. We also collected data on the brewing industry to try to develop the work of Berry and Nix. Methods of non-parametric regression were tried with some success. But as a result of the survey reported in Chapter 6, it was decided to not to pursue them, as they were computationally complex.

7.7 Using regression as a substantive AP

Some of the first works about the effectiveness of different APs were published at the end of the 1970s, including the work of Stringer (1975), Albrecht and McKeown (1976), Kinney (1978), Kaplan (1978 and 1983) and Kinney (1979). They used regression models to predict the value of a particular account area such as income.

Kaplan's (1978) work is typical of the research; he developed a regression model using sales as the explanatory (independent) variable in order to estimate the 'cost of goods sold', or 'accounts receivable'. The model was based on data from a 36-month period. For the audited year, the twelve monthly figures were predicted using the model. He did this for different line items in the financial statements. Only the income figure was achieved with 'reasonable accuracy'; and it was only within 10% of the account value!

Three issues are highlighted by this example. First, the role of the outcome-value was a little different from usual. In this investigation Kaplan was measuring which of the APs got closest to the book-value. This is in contrast with the auditor's usual strategy: which is to set an outcome-value and then to measure the closeness of the prediction by the outcome-rule $|\hat{Y}_T - Y_B| \leq C$. However for the experiment the approach provides information about which AP got closest. The second issue is the nature of the APs used and the third is whether it will identify material error. These issues are discussed here and addressed in the experiments reported in Chapter 8.

The work of Knechel (1988a) will be used as an example of the work of other researchers. It raises a number of issues, including the use of outcome-rules, and the simulation methods used by the researchers and the evaluation program. In Knechel's experiment the account area was income. He tested APs, based on regression models, to see if they would detect different patterns of errors seeded into the accounting data. The investigation-rule was based on a simple linear regression model. There were seven outcome-rules, which are described in Table 7.6.

The accounting data used by Knechel was simulated using the rules described below. In this process Knechel was creating perfect data, in the sense that the explanatory variable was increasing in steps that were only slowly changing and then always in a positive direction. Also there was a high correlation between the explanatory and response variables. He used the following functions to generate the data. An explanatory variable, z_t, was generated such that $z_t \sim N(\mu_t, \sigma(z))$: where $\mu_t = \mu_o + \sum_{i=0}^{t-1} (g\mu_i)$; and where $g = 0.005$, $\mu_o = 100$, $\sigma(z) = 25.00$, and $t = 1{:}48$. The response variable (y_t) was generated using the relationship $y_t = 25 + 5z_t + e$: where $e \sim N(0, s^2)$ and s is the standard error, which was related to the materiality by

$s = \Phi \times$ materiality and Φ was 0.25 for low variability of the data and 0.75 for high variability.

Knechel (1988a) claimed that *'The independent* (explanatory) *variable was generated so that the regression model had a known standard error and an expected* R^2 *of 0.95'*. It is not entirely clear from the published work how this was done, although Knechel claimed it was *'comparable to that* (the method) *used by Kinney and Salamon (1982)'*. This is not discussed here, because the issue arising from Knechel's work does not hinge on this but on the method of the evaluation of the results.

From each data set of 48 pairs of observations, the first 36 pairs were used as the base period, and the last 12 were taken as observations for the period to be audited. The data from the base period were utilized to build the AP model, which was used to predict the monthly values in the audited period. The base period data were assumed to be error free. There is some logic to this, as it is assumed to be data that had been audited and thus attested by the auditor to be free of material error! Although, if there was an ongoing, undetected error in the audited data, then that may affect the ability of the AP to detect a similar pattern of error in the unaudited data. The model would have that error built into its expectation-value.

Three levels of annual error (E_T) at 0%, 2% and 4% were seeded into the 12 months of data from the audited period. For each of these there were four monthly seeding patterns. Knechel, for each set of data, seeded three levels of error in four different patterns. The patterns are shown in Table 7.5.

	Error seeded at random in:	**No error in:**
Case One.	1 month with E_T	11 months
Case Two.	3 months with E_T / 3	9 months
Case Three.	6 months with E_T / 6	6 months
Case Four.	12 months with E_T / 12	0 months

Table 7.5: The different error patterns that where seeded into the audited year.

Although Knechel (1988a) claimed to have 12 combinations of error and error distribution, he really had 9. One with no error and two other levels of error (2% and 4%) spread in four different patterns. Thus his 1,200 unique time-series was really 900. It was 100 time series seeded with 9 patterns of error. This is replicated for the two

conditions $\Phi = 0.25$ or 0.75. Thus he had in total 1800 (i.e. 900×2) unique time series and not the implied number of 2400 (i.e. 1200×2).

For each set of data, the parameters of the expectation model were estimated from the base period (the first 36 months). The expectation models were then used to calculate the expected monthly balances for the audit period. Knechel (1988a) used the outcome-rules shown in the Table 7.6.

	Name given by Knechel	**Outcome-rule**
1	Annual Materiality Rule	Investigate all months if the difference between the annual book value (Y_B) and Upper Confidence Limit ($\hat{Y}_{Tu}$), based on the sum of all twelve months data, exceeds materiality (M). That is if $\lvert\hat{Y}_{Tu} - Y_B\rvert > M$.
2	Monthly Materiality Rule	Investigate any monthly total (Y_b) for which the Upper Confidence Limit for that month ($\hat{Y}_u$) exceeds M/12. That is if $\lvert\hat{Y}_u - Y_b\rvert > M/12$.
3	Combination of [1] and [2].	Used [1], to identify if $\lvert\hat{Y}_{Tu} - Y_B\rvert > M$ and if so apply [2].
4	Rule based on the STAR package	Investigate any month for which $\lvert\hat{Y}_u - Y_b\rvert > M/12$ exceeds M/k, where k=1 for month 1 and so on.
5	Kinney Residual Rule	If $\lvert\hat{Y}_{Tu} - Y_B\rvert > M$, investigate all months with outliers, which are defined to be those with a residual greater than 1.96 times the standard deviation. Investigate first those with the largest residual.
6	Kinney/Salamon/ Uecker rule one	If $\lvert\hat{Y}_{Tu} - Y_B\rvert > M$, investigate all months where prediction deviation exceeds 1.64 times the standard deviation.
7	Kinney/Salamon/ Uecker rule two	If $\lvert\hat{Y}_{Tu} - Y_B\rvert > M$, investigate all months with outliers, which are defined as all months where prediction deviation exceeds 1.96 times the standard deviation.

Table 7.6: The outcome-rules used by Knechel (1988a).

These were typical of the outcome-rules used by others. All these rules use an upper confidence level calculated from a regression model for a specified level of confidence. But Knechel (1988a) did not apply these rules in a simple way: that is to report the results of individual APs directly. He argued that while auditors are interested in information about the number of times an AP would correctly signal error, or fail to signal error, this was not their main concern. They are more interested in whether there is a reduction in other forms of substantive testing. APs are considered cheaper to use

than other substantive tests and so auditors are keen to apply them. In Knechel's experiment he set up rules for calculating the sample reduction using a form of the 'Audit Risk Model' (see section 2.5). The criteria set by Knechel was that the combination of the risks associated with sampling and APs did not exceed 0.05.

An example of the results reported by Knechel (1988a) are sample reductions of up to 31% resulting from the application of an AP with an outcome-rule using a materiality of 2%, and when the probability for the risk associated with the regression results was 0.36. Knechel defined the risk of the AP failing to identify a material error to be the probability associated with the prediction value confidence interval. The reduction in the sample sizes was reported to depend on both the outcome-rule and the distribution of the error, although in terms of sample reduction the outcome-rule has more effect than the distribution of error. If the detection risk associated with the regression is changed to 0.05, the sample sizes are reduced by up to 87%. Knechel (1988a) reported that: *'An important conclusion from this study is that preliminary analytical review techniques can increase audit efficiency (via smaller sample sizes for detail testing) and result in increased audit effectiveness (via lower detection risks)'*. Thus Knechel was claiming some positive impact from the regression based APs.

Knechel's approach, measuring the reduction in sampling, attempts to address a major concern of auditors, the reduction of costs. The mechanism for reducing other tests is the Audit Risk Model, a brief explanation of which was given in section 2.5. As we described in Chapter 2 there are doubts about the validity of that model. Given this uncertainty about the Audit Risk Model, the use of a measure based on it to demonstrate the 'effectiveness' of an AP by measuring how much it reduces other work, such as sampling, may provide false indictors. The use of the Audit Risk Model can distract from the matter of primary concern: do APs predict correctly the presence or absence of error? This is important, because many researchers have measured the effectiveness of an AP by the reduction in another substantive test, such as sampling, rather than reporting directly whether the AP detected error. One example is from the work of Kinney (1979), who demonstrated that there could be a reduction in the size of samples of over 33%. This is consistent with Knechel's results, although it is based on the same, potentially, flawed Audit Risk Model process. The only valid method for

such reductions in the sample sizes would be if the APs support the hypothesis that there is no material error.

Ameen (1989) and Dzeng (1991) also used the idea of combining the assurance arising from APs and tests of detail as a method of measuring the efficiency of the AP. It is this area of the experiments by Knechel (1988a) and the others, the evaluation of the effectiveness of APs using the Audit Risk Model, which could be seen as a shortcoming in their work.

It is not clear whether the reduction argument of Knechel (1988a), Kinney (1979) and the others reflects what auditors do in practice. The Auditing Standards allow such a strategy of using APs to replace other substantive work. They say that: *'Regardless of the assessed levels of inherent and control risks, auditors should perform some substantive procedures for financial statement assertions of material account balances and transaction classes' (SAS 300.8).* They go on to stress that APs may be used on their own to provide the substantive evidence. *'The assessed levels of inherent and control risks cannot be sufficiently low to eliminate the need for auditors to perform any substantive procedures for material account balances and transaction classes. However, these substantive procedures may comprise only analytical procedures where such procedures provide sufficient appropriate evidence'* (author's underlining). It is therefore clear that APs may be used instead of sampling, as was confirmed by the survey reported in Chapter 6. This finding is supported by the results of a survey by Mulligan and Inkster (1999). They found that while some auditors do retain other substantive procedures as a result of AP work but reduce their extent, 29% of all auditors always use APs to replace other substantive tests completely. So APs are clearly used to replace other substantive tests. Interestingly, Ameen (1989) counselled against this and concluded that: *'Although APs may not be efficient and effective when used in isolation, they can be used to reduce the level of test of details and achieve a more efficient audit without sacrificing effectiveness'.*

In terms of promoting APs to auditors anxious to cut costs, the approach of Knechel and others is understandable: it reflects what auditors wish to do. But it does not provide straightforward evidence of the effectiveness of APs, particularly because of the defects in the Audit Risk Model, which might exaggerate the apparent effectiveness of the AP. In terms of measuring the effectiveness of an AP, or a category of APs, the

measurements directly of false negative or false positive signals (type I and II errors) are the essential criteria. As Loebbecke and Steinbart (1987) stated, a concern about an AP is that it must be capable of reliably demonstrating that error is not present.

Another potential criticism of the work of Knechel (1988a) is the use of data generated from a normal distribution. It is near perfect data where the response variable is increasing all the time in small increments. One of his aims was to try and model the whole audit process around APs, taking account of different error distributions, materiality levels, degrees of reliance on APs and samples. '*With these constraints in mind, a simulation method was selected in order to analyse various APs under a range of conditions that correspond to realistic accounting populations*'- Knechel (1988a, page 75). The simulation of normally distributed accounting data is contrary to this aim, as is the generation of a 95% correlation between the response (y_j) and the explanatory variable(s) (x_j 's).

7.8 An example where simple APs were used and where the measurement of effectiveness used the reduction in other substantive testing as the criterion

It was not only the experiments using regression based APs that used the reduction in other substantive testing to measure the effectiveness of the AP. Knechel (1988b) also published work that year using non-statistical APs. He used twelve APs based on ratio and trend models. The data were generated in the same way as in the Knechel experiment described above. Thus he applied regression type APs and these non-statistical APs to similar data. The conclusions of this research were that although these APs reduced the sample sizes, they should not be relied upon entirely to provide audit assurance. Both of Knechel's pieces of work, the 1988a and the 1988b papers, found that *'APs are most effective at detecting large non-recurring errors rather than small, frequent errors that are cumulatively material'*. This finding is not surprising, and supports a conclusion of Loebbecke and Steinbart (1987). We will return to it in the experiments in Chapter 8.

Knechel again measured the effectiveness of an AP by its ability to reduce sampling, the Audit Risk Model being used to calculate the reduction. Possibly because of this, Knechel got some results that are counter-intuitive. Logically, more false-negative signals from an AP might be expected to lead to a larger value for the sample size 'n'

and consequent unnecessary audit effort. However Knechel reported that: *'Sample size decreases were experienced by all the APs regardless of the previously reported frequency of type I errors'*. That is, even though there were false-negative signals, the sample sizes were still decreasing because of the use of the AP.

7.9 Summary and Conclusions

This chapter has considered a number of experiments that have investigated the effectiveness of some of the APs used to gather substantive evidence. The APs were the frequently used 'simple' methods of comparison, trend and ratio analysis, plus regression analysis. The work of Loebbecke and Steinbart (1987) demonstrated that many of the simple APs are not effective, because they have high frequencies of false-negative and false-positive signals. This was true for outcome-values of 5%, 10%, and especially 15%. One would expect the false-negative signals to reduce as the outcome-value increases. It was true that there was some reduction, but even when the outcome-value was 15%, there was still a high frequency of false-negative signals: *'between 18% and 50% depending on the account area and the AP'* (extracted from results of Loebbecke and Steinbart (1987)). In Loebbecke and Steinbart's work, the performance of particular APs varied between different account areas. In addition when any of the APs was applied to a particular account area, its performance varied from year to year. Thus, as well as performing badly, the APs were not consistent in their performance. While it might be thought that regression type APs would perform better than the simple APs, there is some evidence from Kaplan (1978) that this was not uniformly true. Kaplan reported an experiment using APs based on regression in which the expectation value never got closer than within 10% of the book value.

Many other analysts using regression or time series based APs did not report such problems about the effectiveness of APs. Their experiments were reported in section 7.7 and included the work of Knechel (1988a) and others. They evaluated the effectiveness of the APs by reference to the extent of the reduction of another substantive procedures: sampling. The Audit Risk Model was used to calculate the revised substantive sample sizes. Because of the recognized defects of the Audit Risk Model, this reduction does not necessarily provide support for the effectiveness of such APs. The extent of the reduction may be exaggerated or underestimated. Thus the

evidence from these studies is not conclusive in demonstrating whether APs are effective.

These results imply that many APs do not provide reliable evidence. They certainly do not meet the requirement for reliable evidence that is set out in the Statements of Auditing Standards: *'In order to have a reasonable expectation of detecting error or fraud, auditors obtain sufficient reliable audit evidence that puts appropriate emphasis on external evidence or evidence created by the auditors'* (SAS 110: Fraud and Error). Neither would the evidence meet the expectation of the judge in the Pacific Acceptance v Forsyth and Ors case discussed in Chapter 4, that an auditor would obtain '*sufficient, relevant and reliable evidence*'. The judge also said that '*An auditor pays due regard to the possibility of fraud or error by framing and carrying out his procedures, having in mind the general and particular possibilities that exist, to the intent that if a substantial or material error or fraud has crept into the affairs of the company he has a reasonable expectation that it will be revealed'*. These two statements imply that the auditor's tests should have a reasonable chance of detecting material error if it exists. The Concise Oxford Dictionary (10th Edition) defines 'reliability' as '*to be relied on*' and 'rely' to be '*depend on with full trust*'. The results of the AP should be such that the auditor, and others, can depend on them with full trust: that is they are reliable. There should be a high probability that the procedure would consistently discover material error, should that exist, which would certainly require tighter outcome-values than 10% or 15%.

Thus there are two reasons why further investigations are needed to demonstrate if APs, on their own, really fulfil their purpose at the substantive stage. Firstly, we must demonstrate that the APs have low probabilities of false-negative and false-positive signals. To do that an AP, on it's own, should produce reliable evidence of the existence or absence of material error. The performance of the AP must not be confused with, or disguised by, its combination with some other audit substantive procedures. Secondly we need to investigate the effect of the distributions of the data have on the performance of the APs. Both of these are important since we know from Mulligan and Inkster (1999) that APs are the sole source of substantive evidence in at least 29% of all audits.

In Chapter 8 the results of some new experiments are discussed. A number of different investigation-rules are applied to data modelled as if it was from either of two different government agencies, or was similar to that used by Knechel. This will be done for different levels of outcome-rule and seeded error. The investigation-rules will include simple and regression based rules. The effectiveness of the APs will be measured by the frequency of false-negative and false-positive signals.

Chapter 8: Experiments to test the effectiveness of certain APs

8.1 Introduction

This chapter is about the effectiveness of APs when used as substantive procedures. As discussed in Chapter 7, one key question about APs is their ability to correctly signal the presence or absence of error when applied to a line item of a set of financial statements. This chapter reports the investigations as to whether particular investigation-rules (APs) effectively and efficiently signal error. Whether they do is important since APs frequently replace sample based substantive tests.

The Dictionary defines effective to be '*producing a desired or intended result*' (Concise Oxford Dictionary: Tenth Edition 1999), which will be interpreted to be a low frequency of, both, false-positive signals and false-negative signals. The aim was to test the effectiveness of the procedures that, according to the surveys, were frequently used by auditors. So simple procedures were chosen. That is APs with simple investigation-rules, such as those that use comparison type methods, simple trend analysis, ratios, bivariate regression and multivariate regression.

In investigating whether APs with such investigation-rules have low probabilities of false-negative and false-positive signals, the effect of the **error rate is** considered, as well as the impact of the **pattern of error** (for monthly outcome-rules): ranging from errors that occur in large one off amounts, to those which are more evenly distributed through the audited period. This is an issue since some researchers have indicated that the pattern of the error across the audited year influences the ability of APs to detect error. Another variable is the outcome-rule and particularly the **outcome-value**, C, used within that rule. So the performance of the AP will be tested over a range of outcome-values.

In the past much of the assessment of the effectiveness of an AP has used, as a measure, the reduction in other substantive tests resulting from the application of the AP. As was argued in Chapter 7, such a measure is misleading for two reasons. First the methodology for the reduction calculation, the Audit Risk Model, is flawed. Second, it does not reflect what auditors currently do. Typically they replace other substantive tests, such as

sampling, by an AP. So we measured effectiveness of an AP by the frequency of false-negative and false-positive signals of APs. The key variables relevant to the conduct of the experiments are discussed in the next section.

8.2 Issues that are relevant to the investigation of the effectiveness of APs

8.2.1 Precision, or accuracy, of the prediction

Any prediction based on estimation has a precision (or more correctly imprecision) associated with it. If it is a statistical procedure, then a confidence interval can be associated with the expectation-value. In Figure 8.1 the green line shows the book-value (Y_B), the black line shows the true value (Y_T), the red line ($\hat{Y}_T$) shows the expectation value and the blue lines show the lower and upper limits ($\hat{Y}_L$ and $\hat{Y}_U$) of the error bound.

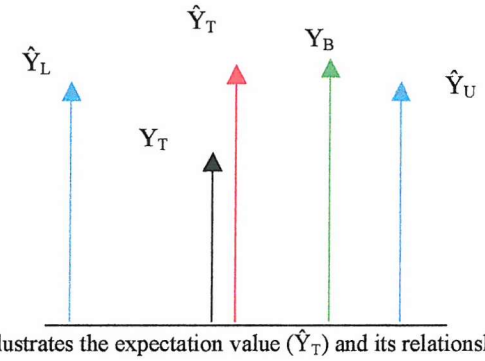


Figure 8.1: This illustrates the expectation value ($\hat{Y}_T$) and its relationship to the book-value Y_B and the true value Y_T.

There are two issues for the closeness of the estimate to the book-value. They are: -

[1] If Y_B is within ($\hat{Y}_L$, $\hat{Y}_U$), the error band for $\hat{Y}_T$, then statistically Y_B is consistent with Y_T. If Y_B is not within ($\hat{Y}_L$, $\hat{Y}_U$), then Y_B is not statistically consistent with Y_T.

[2] If $|\hat{Y}_T - Y_B|$ is large (i.e. material), then Y_B is not auditing consistent with Y_T. That is, there is a possibility of material error. (Auditing consistent means that $|\hat{Y}_T - Y_B|$ is less than some significant amount such as materiality or the outcome-value).

As an example of how these two concepts may not be in agreement, suppose $Y_B > \hat{Y}_T$ and $|\hat{Y}_T - Y_B|$ > material error, but that $Y_B < \hat{Y}_U$. Is Y_B now inconsistent with Y_T? In this case the value is statistically consistent, but not auditing consistent. A contra case might be

where $\hat{Y}_T > \hat{Y}_U$ and $|\hat{Y}_T - Y_B| <$ material error. The apparent contradiction arises because the data is inadequate for drawing both conclusions.

In the survey described in Chapter 6, only one of the firms interviewed considered the error bounds associated with their expectation values. That firm uses a computer package based on the use of regression. In order to reproduce the way the majority of auditors apply their APs, it was decided to ignore the precision associated with the expectation-values. This decision was also based on the short runs of data used by auditors, and the difficulty in providing objective estimates of precision for the simple rules. Therefore, in order to provide consistency between the regression type investigation-rules and the simple rules this thesis will concentrate on the auditing consistency criterion. This issue is taken up again at the end of this Chapter.

8.2.2 The Variables underlying the application of an AP

In setting up the experiments, five variables were identified as being important. They were:

- the investigation-rules,
- the outcome-rule, including the outcome-values,
- the data,
- the level of error, and
- the error pattern.

The purpose in identifying these was to ensure that all the relevant variables in the performance of an AP were considered. It was not to identify which variable was important, or which combinations of variables were important, as is done in a factorial design. The variables are described in more detail in rest of this section.

8.2.2.1 Variable One: the investigation rule

The first variable to be considered is the form of the investigation-rule that is to be applied. The investigation-rules used in the experiments described in this chapter are shown below in Table 8.2. They can be categorised into three classes: bivariate regression (rule 1.1), multivariate regression (rule 1.2) and simple (rules 1.3 to 1.9).

	Description of the investigation-rule	Notes
	Bivariate regression based rule	
1.1	Regression with one explanatory variable	
	Multivariate regression based rules	
1.2	Multivariate regression with three explanatory variables	
1.3	**Comparison** Compare to last year	
1.4 1.5	**Prediction using percentage change in previous years** Year 4 = Year 3 × (1 + % change of years 3 and 2) Year 4 = Year 3 × (Average of % change of years 3 and 2 and years 2 and 1)	
1.6 1.7 1.8	**Prediction using average change in previous years** Year 4 = Year 3 + (Year 2 – Year 1) Year 4 = Year 3 + {(Year 3 – Year 2) + (Year 2– Year 1}/2 Year 4 = Year 3 + {2× (Year 3 – Year 2) + 1× (Year 2– Year 1}/3	Change in previous year Average change in previous years Weighted Average
1.9	**Prediction using average of previous years** Year 4 = {(Year 1 + Year 2 + Year 3)/3}	Simple average

Table 8.2: The Investigation Rules

Regression based investigation-rules were chosen since regression is a well understood procedure and such investigation-rules are similar to those applied by Knechel (1988a) and by Kinney (1979) in their research and in the audit work conducted by Deloitte Touche using their STAR program. Typically, monthly data is used in such procedures, with monthly outcome-rules. This was not to copy the application of such rules, rather the purpose was to identify whether if such rules are used with outcome-rules that directly measure the result of the AP, then the prediction of the effectiveness of the AP will be different to the earlier experiments.

The simple investigation-rules were chosen because they are typical of the procedures that auditors most frequently use. This was demonstrated by the surveys described in Chapters 5 and 6. The chosen procedures all used annual data and are similar to those applied by Loebbecke and Steinbart (1987) in their work, so it will be interesting to see if they produce similar results. The first procedure was comparison: the investigation-rule was to compare the audited value that of last year, and had an associated outcome-rule of 'signal that there is no error if the change between the years was less than the outcome value'. The other simple investigation-rules made use of the book-values for earlier years to build an

expectation-value. Suppose that Y_4 is the book-value for the year being audited and is the total paybill for that year, and that Y_1, Y_2 and Y_3 are equivalent values for the prior years. An investigation-rule might be a prediction based on change in previous years, one such example is $\hat{Y}_4 = Y_3 \times (1 + (Y_3 - Y_2)/Y_2))$.

The other simple investigation-rule was similar, but used the average change in previous years. For example $\hat{Y}_4 = Y_3 + \{a\times(Y_3 - Y_2) + b\times(Y_2 - Y_1)\} / (a + b)$, if $a = b = 1$ this reduces to$\hat{Y}_4 = Y_3 + \{(Y_3 - Y_1)\} / 2$. If $a = 2$ and $b = 1$, then it becomes a weighted average of the yearly differences: $\hat{Y}_4 = Y_3 + \{2\times(Y_3 - Y_2) + (Y_2 - Y_1)\} / (3)$. A variation might be a prediction based on a weighted average of previous years. For example $\hat{Y}_4 = (a\times Y_3 + b\times Y_2 + c\times Y_1)/(a + b + c)$. If $a = b = c = 1$, we have a simple average. Some models give more weight to recent years: for example $a = 3$, $b = 2$ and $c = 1$.

Which of the rules is the appropriate one to apply depends on the data. For example, when considering the simple investigation-rules, if the data are steadily increasing by amounts that do not vary in proportion to the size of the values, then rules 6 to 9 might be best. On the other hand if the data are increasing in a way that is proportional to the values then rules 4 or 5 might be better models.

8.2.2.2 Variable 2: outcome-rules and outcome-values

An outcome-rule of the type $|\hat{Y}_T - Y_B| \leq C$ was used for all the experiments. For investigation-rules 1 – 3, the rule was based on the monthly totals, and for investigation-rules 4 – 9, the rule was based on annual totals. For the experiments described here, there are a number of different criteria that could be used to put a value on that outcome-rule. One is materiality – described in Chapter 2. However materiality is not applied directly. Rather a range of values is used, which could be judged to be significant values and that also reflect the range of values auditors use when considering materiality. Auditors may think of the significant values in terms of proportions or absolute values. For the experiments the outcome-values were set at 1% to 10% of the total book-value being audited: $\sum_{i=37}^{48} y_i$ for the annual rules and of y_i for the monthly rules.

8.2.2.3 Variable 3: the data

A key element of any AP is the data. As explained in Chapter 3, the data used by the auditor might be from the financial systems or may be a combination of that and data from other systems. Those systems may be parts of the audited organisation, or external to it.

To investigate whether APs (investigation-rule) provide high frequencies of false signals it is necessary to find a data set that is well behaved. By that is meant that any increase or decrease in one variable is matched by equivalent changes in other variables and such changes are structured and not prone to large fluctuations. If the AP (investigation-rule) does not perform well, that is it has a potential for false signals, then that will provide strong evidence of a potential flaw in such rules. It will also provide evidence to contradict the earlier experiments whose methodology to measure success was a reduction of other tests. If the APs do well then further experiments may be required.

For the simulations, three sources were used to create the populations. The first two sets of populations were modelled on data from two different Government Agencies: they were the Paymaster Generals Office (PGO) and the Land Registry (LR). These Agencies were chosen because they were audited by the NAO and it was therefore possible to obtain direct evidence about the quality of the data through the work of the NAO. The data from Government Agencies consisted of 48 observations: 48 sets of monthly pay bills from the accounts branch and associated data on staff numbers from a policy branch. In the public sector, the pay area is a well-controlled set of financial and staff information. Management set strict budgets, and staff would soon complain if they were underpaid. Consequently we can assume that it is largely error free: there are management controls that should detect material overpayments by the systems and most underpayments will be the subject of staff action. That assumption is supported by the work of the Internal Auditors of the organisations and the external auditor, the National Audit Office (NAO), both of whom extensively checked the systems and the data.

The data from the Government Agencies contained two features that needed to be adjusted. Since the pay data relates to four different financial years, the payment data needs to be adjusted for the effect of pay inflation to put the data from all years into current values.

Fortunately, pay rises are annual, so the data can easily be put onto a common basis. Pay rises are documented and formally approved by senior management and therefore the pay data can easily and legitimately be adjusted to take account of the effects of these changes. For the experiments, the salary payment data was up-rated so that all the financial data has the same economic value as the data for the year being audited.

The second feature is the timing of the pay rises, a consequence of which is that the pay in a particular month may not relate to the number of staff being paid for that month. This will be particularly significant if pay arrears affect a large proportion of, or all of, the staff. This issue regularly affects the public sector. Pay awards take effect at the start of the financial year, normally on the 1 April. However, because of the nature of large organisations, and sometimes for political reasons, negotiations about pay may last for many months with the results of an award being agreed several months after the beginning of April. The award would then be backdated to the 1 April. Because of this, there may be one particular month where the salary charge is inflated by the effect of the pay arrears, and this month may vary from year to year.

The impact of these two features and how the adjustments were made was explained in Annex A of Chapter 3. From the point of view of the integrity of the audit, such amendments are acceptable provided they are done before any work is started to calculate the expectation value, otherwise this could bias the results. This is an example of why understanding the business is crucial, so that appropriate adjustments can be made and authenticated. In a sense such adjustments are an essential part of the overall AP process.

The data from the agencies reflects other situations. For example the effect of pay drift caused by the incremental pay rises of staff as they progress up pay spines. These were typical of the data that was collected in the early to late 1900's. There is another feature that the data for these experiments does not contain. It is the one off payments to individuals. These are adjustments that relate to all the staff, in respect of pay arrears, pay advances, and repayments. While we have adjusted for the effects of the annual pay awards, and the effects of pay inflation caused by those pay awards, there are other figures in the data underlying the financial statements that are changed by amounts that are not so

easy to establish. The auditor usually does not, or cannot, or should not change these. A figure that does not fit the pattern may just be a random but correct amount, or it may be an error. While we built values that affected the average of these fluctuations, the data does not reflect the extremes. These payments are relatively small in terms of the overall monthly paybill.

The process of simulating the data for the audit populations that are based on the Agencies is shown in the flow diagram in Figure 8.3 below. The file names of the data are shown to enable cross-reference to be made to the programs used to do the simulations.

Using the data from a Government Agency as a prototype, sets of data are constructed that are typical of such organisations. Based on the key parameters of the Agency's data (the mean, standard deviation, minimum and maximum), the suite of programs simulates a set of data. Before being used as a basis for the simulations, the Agency data had been adjusted to current values and for the effects of any back pay arising from the pay awards.

In brief the simulation works in the following way. In order to simulate a set of pay data that has the same characteristics as one of the Agencies we used a suite of programs written in S-Plus. Basically the programs use the parameters of the real data. Suppose there are two explanatory variables representing different staff grades (x_{1i}, x_{2i}) and the response variable (pay) is y_i, for i = 1:n. First the minimum and maximum of the staff variables are calculated. Then using a linear regression model, the parameters a, b_1 and b_2 and the standard deviation (s) are estimated from the Agency data. A new set of data is calculated by $y_i = a + b_1x_{1i} + b_2x_{2i} + N(0, s)$, where i = 1:n and x_{1i} and x_{2i} are uniformly distributed in the range (minimum, maximum). The number of sets of observations (n) is 48 in these experiments.

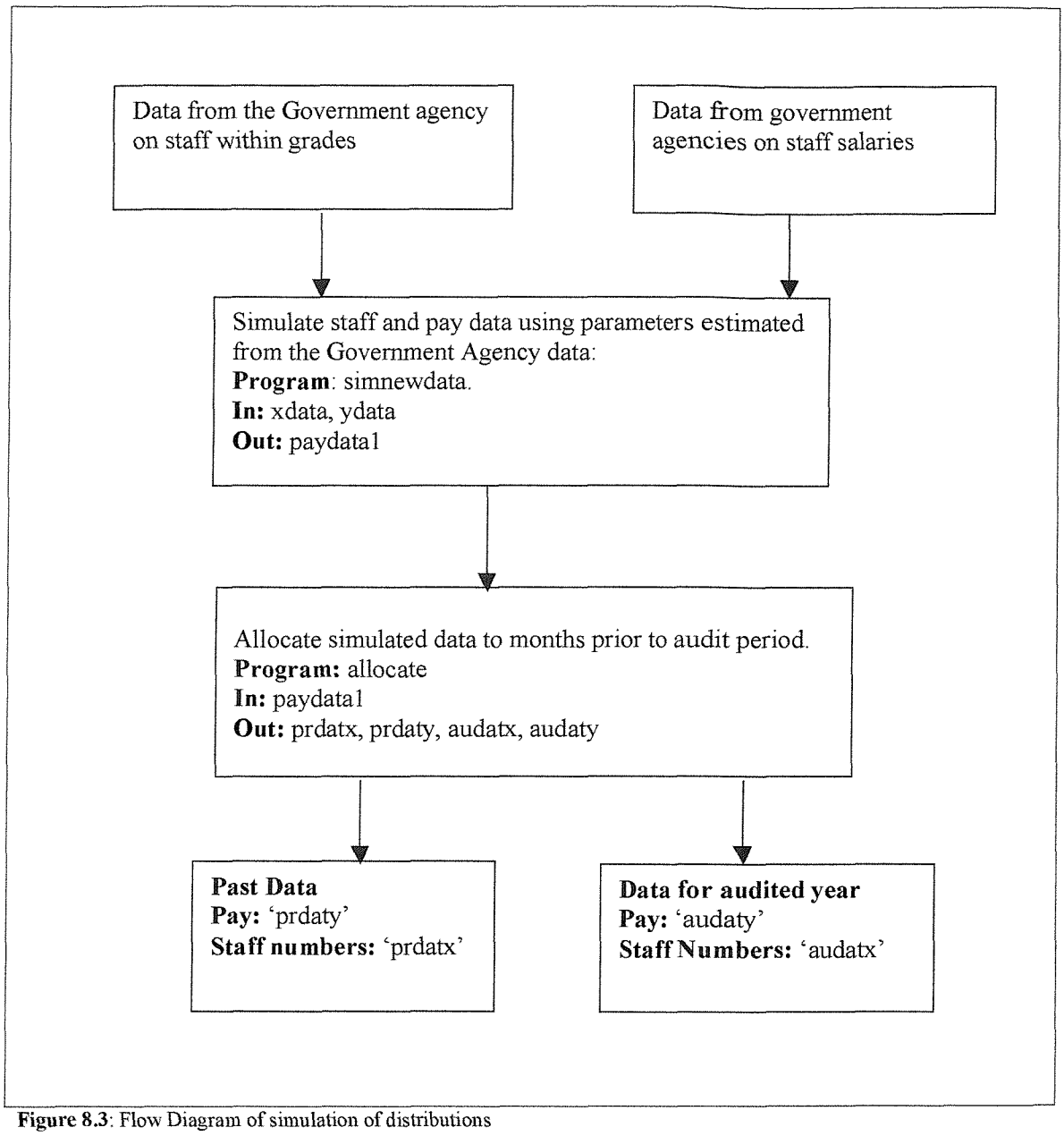


Figure 8.3: Flow Diagram of simulation of distributions

Each new set of simulated data is then allocated between the data that is prior to the audit and the data representing the audited period. For example, if the simulated data were based on one of the Agencies, there would be 48 monthly pay charges to the financial statements and details of the numbers of staff for each of those months. Each set of simulated data is then split into the 36 months of data prior to the audited year and 12 months data for the

audited year. In the first set of data the prior data are in the files 'prdaty' and 'prdatx' and are:

$y_1, \quad y_2, \quad \ldots \quad y_{I+1}, \quad \ldots \quad y_{36} \quad \ldots$ the response variable, and

$x_1, \quad x_2, \quad \ldots \quad x_{i+1}, \quad \ldots \quad x_{36} \quad \ldots$ the explanatory variable

The response and explanatory variables for the year being audited are observations 37 to 48. These are in the files 'audaty' and 'audatx', and are:

$y_{37}, \quad \ldots \quad y_{43}, \quad \ldots \quad y_{48} \quad \ldots$ the response variable, and

$x_{37}, \quad \ldots \quad x_{43}, \quad \ldots \quad x_{48} \quad \ldots$ the explanatory variable

The second Agency yielded a more detailed set of data. They are summarised into two grade groups in the illustration shown below. The pay-bill is the response variable and the number of staff, the explanatory variables. For the prior data we have:

$y_1, \quad y_2, \quad \ldots \quad y_{i+1}, \quad \ldots \quad y_{36} \quad \ldots$ response variable, and

$x_{1,1}, \quad x_{1,2}, \quad \ldots \quad x_{1,i+1}, \quad \ldots \quad x_{1,36} \quad \ldots$ explanatory variable one – senior staff, and

$x_{2,1}, \quad x_{2,2}, \quad \ldots \quad x_{2,i+1}, \quad \ldots \quad x_{2,36} \quad \ldots$ explanatory variable two - junior grades

The response and explanatory variables for the year being audited would be observations 37 to 48, that is:

$y_{37}, \quad \ldots \quad y_{43}, \quad \ldots \quad y_{48} \quad \ldots$ response variable, and

$x_{1,37}, \quad \ldots \quad x_{1,43}, \quad \ldots \quad x_{1,48}$ explanatory variable one – senior staff, and

$x_{2,37}, \quad \ldots \quad x_{2,43} \quad \ldots \quad x_{2,48} \quad \ldots$ explanatory variable two - junior grades

Because the data is from a source that has been extensively audited, the base data is assumed to be error free. The assumption is made for all the simulated data before errors are seeded into the audited data. The seeding of errors into the audited data (y_{37} to y_{48}) will be discussed later in Section 8.2.2.5.

As well as this data from real organisations, one set of the populations was simulated as if it was similar to that used by Knechel (1988a) and was used to replicate a part of his work. The details underlying its generation are in section 7.7.

8.2.2.4 Variable 4: the seeded error

Many auditors have claimed that error rates are low within the data underlying the financial statements. Positive indications of the incidence of errors have come from a number of studies. Kreutzfeldt and Wallace (1986) analysed 260 audits and found that the average number of errors detected and recorded was under 6 per audit. Bell and Knechel (1994) analysed a small number of highly specialised audits (Property and Causality Insurers) and found a low incidence of error in those audits. Even allowing for the caveat that not all errors may be recorded, these examples provide some evidence for a low number of errors in accounting populations.

That audit folklore is taken as a working assumption, but the implications of high error rates will also be considered. The error is seeded into the book-values of the financial data being audited: y_{37}, ... y_{43}, ... y_{48} in the example that is discussed above. The level of error to be seeded was set as a percentage of the total book-value being audited; that is $\sum_{i=37}^{48} y_i$.

This percentage is a convenient way of controlling and explaining the value of the error. The levels of error (E) seeded were from 0% to 15% in increments of 1%, producing 16 levels for variable 4. Another study might choose to look at errors seeded into the explanatory variables.

8.2.2.5 Variable 5: the pattern of error

Other experiments that used monthly outcome-rules, for example Knechel (1988a) and Knechel (1988b), found that the way errors were distributed within an audited period affected the ability of the AP to detect them. The more evenly the error was spread, the harder it was to detect. Conversely large errors, occurring as if all the error was concentrated at one point, were more likely to be detected. To measure the impact of the different error patterns, the error was seeded in four different patterns. These are shown in Table 8.4. The index number is the flag of whether the error is allocated at random to one month in the twelve, or spread uniformly over 3, 6 or 12 months. In the 1, 3 and 6-month cases the seeded months are selected, at random.

Error Pattern	Index Number
All in one of the twelve months	2
Spread over three months	3
Spread over six months	4
Spread over all twelve months	5

Table 8.4: The error pattern seeded into the simulations

For annual outcome-rules the error pattern was not an issue.

8.3 A Description of the Experiments

The objective of the investigations was to indicate the extent to which APs were effective, with different sets of data, for different investigation rules and for different combinations of error, both level and pattern. First the way in which the different variables were considered in the experiments is described. To illustrate how this was done, the implementation of one combination of variables is used as a basis to describe the outline of the experiments. All the experiments were organized in a similar way. The process was implemented using a suite of programs written in S-Plus, a language that was chosen because it offers certain routines, such as regression and random number generation that can be incorporated into any routines / programs that are written.

The rest of this section describes the implementation of the experiments to measure the level of false-negative and false-positive signals for the APs.

8.3.1 The implementation of the experiments

The description of the process underlying the implementation of the experiments is summarised in the flowchart in Figure 8.5 on page 198. Briefly the investigation-rule is selected [Step 1], and a data set chosen as the basis for the simulated data. In the example described here they were, respectively, simple regression and the data from one of the Government Agencies. Then the experiments follow the sequence of the flow chart.

In step 2 we simulate a population to which the AP is applied. For each iteration we generate 48 sets of observations (pay and staff data) to represent the monthly data for the four-year period. This follows the method described in Section 8.2.2.3.

After the data is generated, the next step (step 3) is to sort it into the four subpopulations: the first 36 sets of variables are the prior data which are divided into explanatory (xdata) and response variables (ydata), and the final 12 are the audited data which are divided into explanatory (audatx) and response variables (audaty). At this stage the data in audaty represents Y_{Ti}, the "true" book-values for each month.

In step 4, a model is built that represents the investigation-rule. The prior data (xdata and ydata) are used to build the model. Then the expectation values ($\hat{Y}_{Ti}$, where i = 1...12 for monthly data) are estimated using the explanatory variable(s) (audatx) from the audited period.

Steps 5 and 6 work in conjunction with each other, and are parallel to step 4. The initial step in the process of seeding the error, step 5, is the selection of the error patterns that will be used for this part of the experiment. That pattern is then fixed until we have evaluated all combinations of error and outcome-value.

Step 6 is to seed the error into 'audaty', to produce the Y_{Bi} values. An assumption is made that the error is in the financial statements, and the aim of the AP is to signal when there is an imbalance (error signalled) between the financial statements and the underlying data. The error is therefore seeded within the values being estimated: $Y_1, Y_2, .., Y_{12}$. This is done first for E = 0 (0% error), and then for E = 1% to 15%.

Step 7 is to compare $\hat{Y}_T$ to Y_B using outcome rule, followed by Step 8 to record the results of whether $|\hat{Y}_{TE} - Y_{BE}| \leq C$ for C = 1% to 10%. A score of "1" is recorded if $|\hat{Y}_{TE} - Y_{BE}| \leq C$ and "0" otherwise. For each time round the loop the scores are recorded into an output matrix, which is then updated with each successive evaluation.

Thus if a score of "1" is recorded that is added into the total for each combination of outcome-value and seeded error value. At the end of the process the proportion of times that $|\hat{Y}_{TE} - Y_{BE}| \leq C$ can be calculated by dividing through by the number of simulations in the trial. The table of observed frequencies is computed from this matrix.

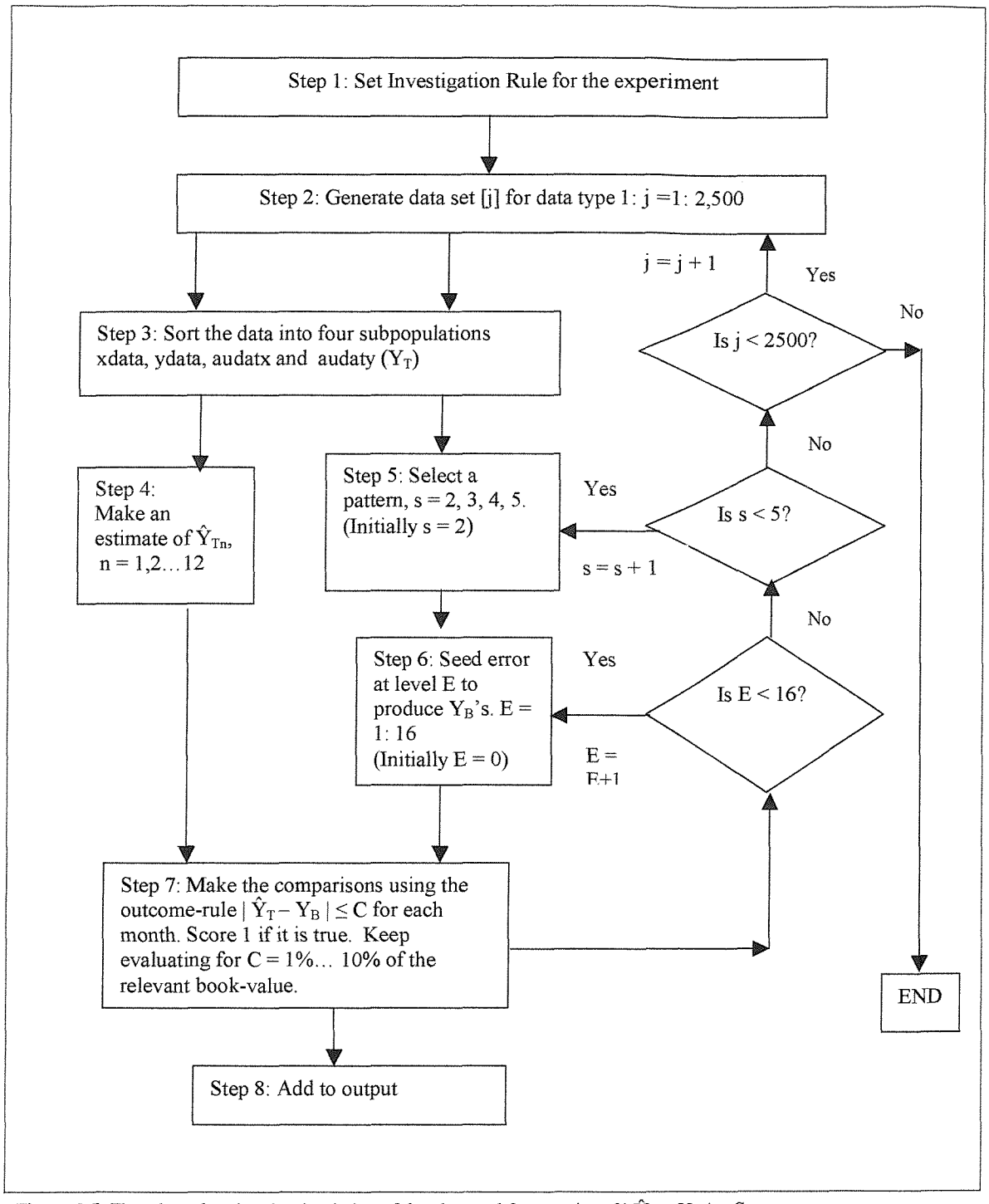


Figure 8.5: Flowchart showing the simulation of the observed frequencies of $|\hat{Y}_T - Y_B| \leq C$.

An example of the form of the final output is in Table 8.7 on page 200. The outcome-rule, $|\hat{Y}_{TE} - Y_{BE}| \leq C$, is evaluated on a monthly basis, recording the number of times this was true [t] for each month, in which case t = 1. This result is calculated for each of the 12

months, so there are a series of 0's and 1's. The average frequency for the year is then given by $\sum_{i=1}^{12} t_i/12$. These are then updated by successive simulations (2,500 in these trials).

The process is then repeated for all patterns of error (error level and pattern). Once that is done, the whole of the above is repeated for the next set of simulated data (from the same basic source of data, be it one of the agencies or the other data) until j = 2,500 when the process is stopped and a table is produced of the outcomes.

Thus a frequency has been computed for observing whether $|\hat{Y}_{TE} - Y_{BE}| \leq C$, for each combination of variables: variable 2 (data at one level), variable 4 (error rate at 16 levels) × variable 5 (error pattern at 4 levels) evaluated over variable 2 (outcome-values at 10 levels). The total number of simulations for experiment 1 is 1,600,000. That is (1 × 1× 4× 15× 10× 2500) for E = 1% to 15%, plus (1×1×4× 1× 10 ×2,500) for the case when the no error (E = 0 in Figure 8.5) is seeded and the there is therefore no pattern of error. The combination of variables is summarised in Table 8.6.

Variable	Level	Notes
Investigation rule	1	
Data	Land Registry	
Error pattern	4	
Error level	16	The 0% rate not applied by all of variables
Outcome rule	10	The monthly outcome-rule

Table 8.6: This shows the variables used for a particular experiment.

We now have the frequency of observing that $|\hat{Y}_T - Y_B| \leq C$ for all combination of error and outcome-value. An example of the form of the output is in Table 8.7 on page 200. Each cell records the relative frequencies of observing if $|\hat{Y}_T - Y_B| \leq C$ for each combination of outcome-value (C) and seeded error (E).

The important question is whether these frequencies are close to those that would have occurred if the comparison were to the true value (Y_T), rather than $\hat{Y}_T$. That is $|Y_T - Y_B| \leq C$. This would present the value that a 'perfect' AP would produce. We calculated such frequencies using the same rationale as that to calculate the frequency of observing $|\hat{Y}_T - Y_B| \leq C$ (Figure 8.5). The difference is that instead of estimating $\hat{Y}_T$ for each month, we

take the value of Y_B before seeding the error as the estimate of the true value Y_T. We then seed the error as before to create the new value for Y_B, that is Y^1_T. The outcome-rule for this is $|Y^1_T - Y_B| \leq C$.

Outcome Values	Error Levels (Over 3 months)															
	0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	11%	12%	13%	14%	15%
One %	0.16	0.14	0.13	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12
Two %	0.31	0.29	0.25	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23
Three %	0.45	0.42	0.37	0.35	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34
Four %	0.57	0.54	0.48	0.44	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43
Five %	0.67	0.64	0.57	0.53	0.51	0.50	0.50	0.50	0.51	0.50	0.50	0.50	0.51	0.51	0.51	0.51
Six %	0.76	0.73	0.66	0.60	0.58	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57
Seven %	0.83	0.80	0.74	0.67	0.63	0.62	0.62	0.62	0.62	0.62	0.62	0.62	0.62	0.62	0.62	0.62
Eight %	0.88	0.86	0.80	0.72	0.68	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66
Nine %	0.92	0.90	0.85	0.77	0.72	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69
Ten %	0.94	0.93	0.89	0.82	0.75	0.72	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71

Table 8.7: This is an example of the output from a run of the program to produce the frequency of observing $|Y^1_T - Y_B| \leq C$.

The Table 8.8 below shows the expected frequencies of $|Y_T - Y_B| \leq C$ for the case when the error is seeded equally into three months. As will be seen, these entries are close to the theoretical expectations. The AP should identify that nine months have no error, and error will be identified in the remaining three months. Hence the probability of observing that there is no error is $\geq (9 \times 1 + 3 \times 0)/12 = 0.75$.

Outcome Values	Error Rates (Over 3 months)															
	0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	11%	12%	13%	14%	15%
One %	1.00	0.75	0.75	0.75	0.74	0.74	0.75	0.75	0.75	0.75	0.76	0.75	0.76	0.74	0.75	0.75
Two %	1.00	0.74	0.76	0.75	0.75	0.75	0.74	0.75	0.75	0.77	0.76	0.75	0.75	0.73	0.74	0.74
Three %	1.00	0.75	0.75	0.76	0.74	0.75	0.74	0.76	0.76	0.75	0.75	0.75	0.74	0.74	0.75	0.76
Four %	1.00	0.89	0.75	0.75	0.76	0.76	0.76	0.76	0.75	0.75	0.73	0.74	0.75	0.74	0.75	0.76
Five %	1.00	1.00	0.75	0.74	0.75	0.75	0.76	0.75	0.74	0.74	0.75	0.76	0.77	0.75	0.74	0.75
Six %	1.00	1.00	0.75	0.76	0.77	0.75	0.77	0.73	0.75	0.76	0.76	0.75	0.75	0.75	0.74	0.73
Seven %	1.00	1.00	0.84	0.75	0.77	0.76	0.74	0.75	0.74	0.75	0.76	0.75	0.74	0.77	0.76	0.74
Eight %	1.00	1.00	0.92	0.76	0.75	0.74	0.75	0.76	0.75	0.75	0.74	0.74	0.75	0.75	0.75	0.74
Nine %	1.00	1.00	0.99	0.75	0.75	0.76	0.75	0.74	0.76	0.74	0.75	0.75	0.74	0.75	0.76	0.76
Ten %	1.00	1.00	1.00	0.82	0.75	0.75	0.76	0.76	0.75	0.73	0.74	0.75	0.76	0.77	0.76	0.76

Table 8.8: This shows the expected frequencies of observing that $|Y_T - Y_B| \leq C$.

Thus we have a table of observed frequencies (probabilities), which can be compared to a table of frequencies that would have occurred if the AP exactly predicted the true book-value. We will call these frequencies p_O and p_E respectively.

For the annual investigation-rules, the data is produced in the same way, but is then summed to produce annual totals. This was done to ensure consistency across all the

experiments. The tables of observed and expected frequencies are then computed in a similar way to those above.

8.3.2 Interpreting the results of the experiments

To illustrate the process of interpreting the results of the experiments, let us consider the case where the seeded error (E) is 0: the results of which are in Table 8.9. There the observed and expected frequencies are set out so we can compare them directly.

No error seeded into audited population		Error = 0%		
		p_O	p_E	Diff
Outcome-Values	One %	0.16	1.00	-0.84
	Two %	0.31	1.00	-0.69
	Three %	0.45	1.00	-0.55
	Four %	0.57	1.00	-0.43
	Five %	0.67	1.00	-0.32
	Six %	0.76	1.00	-0.24
	Seven %	0.83	1.00	-0.17
	Eight %	0.88	1.00	-0.12
	Nine %	0.92	1.00	-0.08
	Ten %	0.94	1.00	-0.06

Table 8.9: The table shows the observed and expected frequencies of $|\hat{Y}_T - Y_B| \leq C\%$ if the error rate is zero.

Take the case where C = 3% of the book-value. If we consider the observed frequency (p_O), the frequency of $|\hat{Y}_T - Y_B| \leq 3\%$ was 0.45. Actually, there was no error in the population, so the AP should have scored '1' in all of the trials. Thus there was a false-negative signal of 0.55 for this combination of error and outcome-value. This is an alarming result, since it implies that the auditor could be undertaking nugatory work in a large number of the applications of an AP.

The situation when error is seeded into the audited population (the Y_i's) is slightly more complex to interpret. For a monthly outcome-rule, this is replicating an audit strategy of searching for each month where the error is greater than the outcome-value for that month. Take as an example the case when the error is seeded into one month of the twelve (Variable 3: s = 2). For the situation when the seeded error is set at 3%, the results are as set out in Table 8.10. Taking the outcome-value to be C = 3%, as before, we have the observed frequency (p_O) = 0.41 while the expected frequency (p_E) was 0.91: a false-negative signal of 0.5.

Error seeded into one month of the twelve		Error seeded at 3%		
		p_O	p_E	Diff
Outcome-Values	One %	0.14	0.91	-0.76
	Two %	0.28	0.91	-0.63
	Three %	0.41	0.91	-0.50
	Four %	0.52	0.91	-0.39
	Five %	0.62	0.92	-0.30
	Six %	0.70	0.91	-0.21
	Seven %	0.76	0.90	-0.14
	Eight %	0.81	0.90	-0.09
	Nine %	0.84	0.91	-0.07
	Ten %	0.87	0.90	-0.04

Table 8.10: The table shows the observed and expected frequencies of $|\hat{Y}_T - Y_B| \leq C\%$ if the error rate is 3%.

Now the general principles have set out of how the results of the simulation are recorded and interpreted, the next step is to look at the interpretation of the full experiment. As was set out at the start of this chapter, the intention is to indicate the extent of the false-negative and false-positive signals. Because of the nature of the results of the trials, a system of colour coding is adopted. It is analogous to some contour maps where different colours signal the relative height above a specified datum level. A similar approach is adopted: the degrees of effectiveness in terms of the relative frequency of false-negative and false-positive signals will be shown by the colours used in the table of results of observing that $|\hat{Y}_T - Y_B| \leq C$. These are shown in Table 8.11.

	Relative frequency of signals that are false resulting from the application of an AP.	**Colour used to indicate whether false-negative or false-positive.**
$p_O < p_E$	Frequency of false-negative signals is ≥ 0.5	Dark Blue
$p_O < p_E$	Frequency of false-negative signals is ≥ 0.2 but < 0.5	Blue
$p_O < p_E$	Frequency of false-negative signals is ≥ 0.1 but < 0.2	Light Blue
$p_O \approx p_E$	Frequency of a false signal is < 0.1	Clear (White)
$p_O > p_E$	Frequency of false-positive signals is ≥ 0.1 but < 0.2	Pink
$p_O > p_E$	Frequency of false-positive signals is ≥ 0.2 but < 0.5	Middle Red
$p_O > p_E$	Frequency of false-positive signals is ≥ 0.5	Red

Table 8.11: This shows the key to the colour coding of the results of the experiments

Although the entry in each cell is the observed frequency that $|\hat{Y}_T - Y_B| \leq C$, the thesis concentrates on the levels of false-signals. In broad terms the interpretation of the colours

is: the darker the colour, red for false-positive and blue for false-negative, the greater the level of false signals and therefore the worse the performance of the investigation-rule.

8.4 The Results of the Experiments

A list of all the experiments that have been carried out are set out in the Table 8.12, and the details of the experiments are described in the rest of this chapter.

Experiment	Variable One	Variable 2	Variable 3	Variable 4	Variable 5
	(Investigation-rule)	(Outcome -rule)	(Data source)	(Error rule)	(Error Pattern)
One	Bivariate Regression	Monthly	LR	All	All
Two	Bivariate Regression	Monthly	PGO	All	All
Three	Multivariate Regression	Monthly	PGO	All	All
Four	Bivariate Regression	Monthly	Similar to Knechel	All	All
Five - Eleven	Various Simple APs	Annual	PGO & LR	Annual	N/A

Table 8.12: This table provides a summary of the variables used in particular experiments

The results are presented in a tabular format which follows that of Table 8.13 on page 205. Each cell in the table shows, for a particular combination of error and outcome-value, the frequency, in the repeated trials, that $|\hat{Y}_T - Y_B| \leq C$. That is the frequency that the difference between the expectation-value ($\hat{Y}_T$) and the book value (Y_B) is less than the relevant outcome-value (C). The values of the outcome-value and error are percentages of the value being audited (Y_B) by the substantive test. The colouring shows the accuracy of the observation as explained in Table 8.11. The darker the colour the worse the performance: that is the higher the frequency of false signals. In each experiment the interpretation of the results is first made for the cases where the error (E) $\leq 5\%$. That is then extended to include all values in the range $0\% \leq E \leq 15\%$.

For each combination of outcome-value and error there were 2,500 simulations to calculate the frequency that $|\hat{Y}_T - Y_B| \leq C$ for the particular investigation-rule that was used in that experiment. In addition there were a further 2,500 simulations for each combination of outcome-value and error to calculate the equivalent frequency if the investigation rule had been perfect.

The results for the regression-based investigation-rules, when used with monthly outcome-rules, are shown in Tables 8.13 to 8.16 (and described in sections 8.5 to 8.8). The first

quarter of each table show the situation when all the error was seeded into one of the twelve months, the month being selected at random. The second quarter of the tables shows the situation when one third of the error was seeded into each of three months, selected at random. Similarly, the third and fourth quarters of the tables show the cases when the error is seeded into six and twelve months respectively.

The consideration of the regression-based rules was followed by a description of the results when simple investigation-rules are used with annual outcome-rules (section 8.9). The investigation-rules that are used are shown in Table 8.17 and the results in Tables 8.18 to 8.24. A brief description of each experiment is given, followed by an interpretation of the result. Later, in Section 8.10, the findings and implications are summarized and discussed.

8.5 Experiment 1: The use of Bivariate Regression with data from an Agency

The results for the first experiment are shown in Table 8.13. The data was simulated to have the same characteristics as the pay data from the Land Registry Agency (LR). An inspection of the table shows that when no error is seeded into the data for the audited period, there are varying levels of false-negative signals, unless $C \geq 9\%$.

The results are now considered when there is seeded error. If $E \leq 2\%$ and if $C \leq 8\%$ there were false-negative signals, for all patterns of error. Also, while the error is in the range $3\% \leq E \leq 5\%$, then depending on the spread of the error across the months, there are varying degrees of false-negative signals for all levels of the outcome-value (C). When the error is seeded into one month, there were false-negative signals while $C < 8\%$. The corresponding figures when the error is spread over three or six months were $C < 7\%$ and $C < 3\%$. If the error was spread evenly over the twelve month period, then there were false-positive signals while $C < E$, and false-negative signals otherwise.

Considering the outcomes of the experiment for all values of $E > 5\%$, it can be seen from Table 8.13 that, generally, the above observations still apply. Overall the investigation-rule performed badly for this data set, giving false-negative and false-positive signals depending on the error spread.

Outcome Values		Error all in one month															
		0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	11%	12%	13%	14%	15%
	One %	0.16	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14
	Two %	0.31	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28
	Three %	0.45	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41
	Four %	0.57	0.53	0.52	0.52	0.52	0.52	0.52	0.52	0.52	0.52	0.52	0.52	0.52	0.52	0.52	0.52
	Five %	0.68	0.63	0.62	0.62	0.62	0.62	0.62	0.62	0.62	0.62	0.62	0.62	0.62	0.62	0.62	0.62
	Six %	0.76	0.71	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70
	Seven %	0.83	0.78	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76
	Eight %	0.88	0.83	0.81	0.81	0.81	0.81	0.81	0.81	0.81	0.81	0.81	0.81	0.81	0.81	0.81	0.81
	Nine %	0.92	0.87	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84
	Ten %	0.95	0.90	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87
		Error spread over three months															
		0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	11%	12%	13%	14%	15%
	One %	0.16	0.14	0.13	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12	0.12
	Two %	0.31	0.29	0.25	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23	0.23
	Three %	0.45	0.42	0.37	0.35	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34	0.34
	Four %	0.57	0.54	0.48	0.44	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43
	Five %	0.67	0.64	0.57	0.53	0.51	0.50	0.50	0.50	0.51	0.50	0.50	0.50	0.51	0.51	0.51	0.51
	Six %	0.76	0.73	0.66	0.60	0.58	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57	0.57
	Seven %	0.83	0.80	0.74	0.67	0.63	0.62	0.62	0.62	0.62	0.62	0.62	0.62	0.62	0.62	0.62	0.62
	Eight %	0.88	0.86	0.80	0.72	0.68	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66
	Nine %	0.92	0.90	0.85	0.77	0.72	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69
	Ten %	0.94	0.93	0.89	0.82	0.75	0.72	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71
		Error spread over six months															
		0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	11%	12%	13%	14%	15%
	One %	0.16	0.15	0.13	0.12	0.10	0.09	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08
	Two %	0.31	0.30	0.27	0.23	0.20	0.18	0.17	0.16	0.16	0.16	0.15	0.15	0.16	0.16	0.15	0.16
	Three %	0.44	0.43	0.39	0.34	0.30	0.26	0.24	0.23	0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.22
	Four %	0.57	0.56	0.51	0.45	0.39	0.35	0.32	0.30	0.29	0.29	0.29	0.29	0.29	0.28	0.28	0.28
	Five %	0.68	0.66	0.62	0.55	0.48	0.43	0.38	0.36	0.35	0.35	0.34	0.34	0.34	0.34	0.34	0.34
	Six %	0.76	0.75	0.71	0.64	0.57	0.50	0.45	0.41	0.40	0.39	0.38	0.38	0.38	0.38	0.38	0.38
	Seven %	0.83	0.82	0.78	0.72	0.65	0.57	0.51	0.46	0.44	0.43	0.42	0.42	0.42	0.41	0.41	0.41
	Eight %	0.88	0.88	0.85	0.79	0.72	0.64	0.57	0.52	0.48	0.46	0.45	0.44	0.44	0.44	0.44	0.44
	Nine %	0.92	0.91	0.89	0.84	0.78	0.70	0.63	0.56	0.52	0.49	0.47	0.46	0.46	0.46	0.46	0.46
	Ten %	0.94	0.94	0.93	0.89	0.83	0.76	0.69	0.62	0.56	0.52	0.50	0.48	0.48	0.47	0.47	0.47
		Error spread over all months															
		0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	11%	12%	13%	14%	15%
	One %	0.16	0.16	0.15	0.13	0.11	0.10	0.08	0.06	0.05	0.04	0.02	0.01	0.01	0.01	0.00	0.00
	Two %	0.31	0.30	0.29	0.26	0.23	0.20	0.16	0.13	0.10	0.07	0.05	0.03	0.02	0.01	0.01	0.01
	Three %	0.45	0.44	0.42	0.39	0.34	0.30	0.25	0.20	0.16	0.12	0.08	0.06	0.04	0.02	0.01	0.01
	Four %	0.57	0.56	0.54	0.50	0.46	0.40	0.34	0.27	0.22	0.17	0.12	0.09	0.06	0.04	0.03	0.02
	Five %	0.67	0.67	0.65	0.61	0.56	0.50	0.43	0.36	0.29	0.23	0.18	0.13	0.09	0.06	0.04	0.03
	Six %	0.76	0.76	0.74	0.70	0.65	0.59	0.52	0.45	0.37	0.30	0.24	0.19	0.14	0.10	0.07	0.05
	Seven %	0.83	0.83	0.81	0.78	0.74	0.68	0.61	0.54	0.46	0.39	0.31	0.25	0.19	0.15	0.11	0.07
	Eight %	0.88	0.88	0.87	0.84	0.81	0.76	0.69	0.63	0.55	0.48	0.40	0.33	0.26	0.21	0.16	0.11
	Nine %	0.92	0.92	0.91	0.89	0.86	0.82	0.77	0.71	0.64	0.57	0.50	0.42	0.34	0.28	0.22	0.16
	Ten %	0.95	0.95	0.95	0.93	0.91	0.87	0.83	0.78	0.72	0.66	0.58	0.51	0.44	0.36	0.29	0.23

Table 8.13: Experiment one with the investigation-rule 1 and data simulated to be like LR.

8.6 Experiment 2: The use of Bivariate Regression with data from another Agency

Table 8.14 shows the results of this experiment. In essence, it is a repeat of Experiment One, but with the data simulated to be similar to that from a different Government Agency: the Paymaster Generals Office (PGO).

The results are similar to experiment one, with variations in detail: generally there were higher levels of false signals. For example, when the error is zero, for all outcome values

less than or equal to 10%, the relative frequency of false-negative signal is greater than 0.1. The comparable figure in experiment one was 8%.

Outcome Values	Error spread over 1 month															
	0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	11%	12%	13%	14%	15%
One %	0.11	0.11	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
Two %	0.23	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21
Three %	0.33	0.31	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30	0.30
Four %	0.43	0.40	0.39	0.39	0.39	0.39	0.40	0.39	0.40	0.39	0.39	0.39	0.40	0.39	0.39	0.39
Five %	0.52	0.49	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48
Six %	0.61	0.57	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56
Seven %	0.68	0.65	0.63	0.63	0.63	0.63	0.63	0.63	0.63	0.63	0.63	0.63	0.63	0.63	0.63	0.62
Eight %	0.75	0.71	0.69	0.69	0.69	0.68	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.68
Nine %	0.80	0.77	0.74	0.74	0.74	0.73	0.74	0.74	0.73	0.73	0.73	0.73	0.74	0.73	0.74	0.73
Ten %	0.85	0.81	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.77	0.78	0.78
	Error spread over 3 months															
	0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	11%	12%	13%	14%	15%
One %	0.11	0.11	0.10	0.09	0.09	0.09	0.08	0.08	0.08	0.09	0.08	0.08	0.08	0.08	0.08	0.08
Two %	0.22	0.21	0.20	0.18	0.17	0.17	0.16	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17
Three %	0.33	0.32	0.29	0.27	0.25	0.25	0.24	0.24	0.25	0.25	0.24	0.25	0.25	0.24	0.25	0.25
Four %	0.43	0.42	0.38	0.35	0.33	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32
Five %	0.52	0.51	0.47	0.44	0.41	0.40	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39	0.39
Six %	0.61	0.59	0.55	0.51	0.48	0.46	0.46	0.45	0.45	0.46	0.45	0.46	0.46	0.45	0.46	0.46
Seven %	0.68	0.67	0.62	0.58	0.54	0.52	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51	0.51
Eight %	0.74	0.73	0.69	0.64	0.60	0.57	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56	0.56
Nine %	0.80	0.79	0.75	0.70	0.65	0.62	0.61	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.60
Ten %	0.84	0.84	0.80	0.75	0.69	0.66	0.64	0.63	0.63	0.63	0.63	0.63	0.63	0.63	0.63	0.63
	Error spread over 6 months															
	0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	11%	12%	13%	14%	15%
One %	0.11	0.11	0.10	0.09	0.09	0.08	0.07	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
Two %	0.22	0.22	0.21	0.19	0.17	0.16	0.14	0.13	0.12	0.12	0.11	0.11	0.11	0.11	0.11	0.11
Three %	0.33	0.32	0.31	0.28	0.26	0.23	0.21	0.19	0.18	0.17	0.17	0.17	0.16	0.16	0.17	0.17
Four %	0.43	0.43	0.41	0.37	0.34	0.31	0.28	0.26	0.24	0.23	0.22	0.22	0.21	0.22	0.22	0.22
Five %	0.52	0.52	0.49	0.46	0.42	0.38	0.35	0.32	0.29	0.28	0.27	0.27	0.26	0.26	0.26	0.26
Six %	0.61	0.60	0.58	0.54	0.50	0.45	0.41	0.38	0.34	0.33	0.32	0.31	0.30	0.30	0.31	0.30
Seven %	0.68	0.68	0.65	0.62	0.57	0.52	0.48	0.44	0.40	0.38	0.36	0.36	0.34	0.34	0.35	0.34
Eight %	0.74	0.74	0.72	0.68	0.64	0.59	0.54	0.49	0.45	0.42	0.40	0.39	0.38	0.37	0.38	0.37
Nine %	0.80	0.80	0.78	0.74	0.70	0.65	0.59	0.54	0.50	0.46	0.44	0.43	0.41	0.40	0.40	0.40
Ten %	0.84	0.84	0.83	0.80	0.75	0.70	0.65	0.60	0.54	0.51	0.48	0.46	0.44	0.43	0.43	0.42
	Error spread over 12 months															
	0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	11%	12%	13%	14%	15%
One %	0.11	0.11	0.10	0.10	0.10	0.09	0.08	0.07	0.06	0.05	0.04	0.03	0.03	0.02	0.02	0.01
Two %	0.23	0.22	0.21	0.20	0.19	0.18	0.17	0.14	0.12	0.10	0.08	0.07	0.06	0.04	0.04	0.03
Three %	0.33	0.33	0.32	0.31	0.29	0.27	0.24	0.22	0.19	0.16	0.13	0.11	0.09	0.07	0.05	0.04
Four %	0.43	0.43	0.42	0.40	0.38	0.35	0.32	0.29	0.25	0.22	0.18	0.15	0.12	0.10	0.08	0.06
Five %	0.53	0.52	0.51	0.50	0.47	0.43	0.39	0.36	0.32	0.28	0.24	0.20	0.16	0.13	0.11	0.08
Six %	0.61	0.61	0.60	0.58	0.55	0.51	0.47	0.43	0.38	0.34	0.30	0.25	0.21	0.17	0.14	0.11
Seven %	0.68	0.68	0.67	0.65	0.63	0.59	0.55	0.50	0.45	0.41	0.36	0.31	0.27	0.22	0.18	0.15
Eight %	0.75	0.75	0.74	0.72	0.69	0.66	0.62	0.58	0.53	0.47	0.42	0.37	0.32	0.28	0.23	0.19
Nine %	0.80	0.81	0.80	0.78	0.76	0.73	0.69	0.64	0.60	0.54	0.49	0.44	0.38	0.33	0.29	0.24
Ten %	0.85	0.85	0.84	0.83	0.81	0.78	0.75	0.71	0.66	0.61	0.56	0.50	0.45	0.40	0.35	0.30

Table 8.14: Experiment Two with the investigation rule 1 and data simulated to be like PGO1

8.7 Experiment 3: Results of experiment using multivariate regression with data that is distributed as if it were from the PGO Agency

This experiment explores the impact of using data that is disaggregated. The data used is from the same area within the PGO Agency as in Experiment 2, but with three explanatory variables rather than one. The total staff numbers that were used in Experiment Two are disaggregated into three different staff grades. The results are shown in Table 8.16.

		Error spread over 1 month															
		0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	11%	12%	13%	14%	15%
Outcome Values	One %	0.29	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26
	Two %	0.54	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49
	Three %	0.72	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66	0.66
	Four %	0.85	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78	0.78
	Five %	0.93	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85
	Six %	0.97	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89
	Seven %	0.99	0.91	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90
	Eight %	1.00	0.92	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91
	Nine %	1.00	0.93	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92
	Ten %	1.00	0.95	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92
		Error spread over 3 months															
		0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	11%	12%	13%	14%	15%
Outcome Values	One %	0.28	0.24	0.21	0.21	0.21	0.21	0.22	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21
	Two %	0.53	0.46	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40	0.40
	Three %	0.72	0.63	0.55	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54	0.54
	Four %	0.85	0.76	0.66	0.64	0.63	0.64	0.64	0.63	0.63	0.63	0.63	0.63	0.63	0.64	0.64	0.63
	Five %	0.92	0.86	0.73	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69
	Six %	0.97	0.93	0.79	0.73	0.72	0.73	0.73	0.73	0.72	0.73	0.73	0.72	0.72	0.72	0.73	0.72
	Seven %	0.99	0.96	0.85	0.75	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74
	Eight %	1.00	0.98	0.89	0.78	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75
	Nine %	1.00	0.99	0.93	0.81	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75
	Ten %	1.00	1.00	0.96	0.84	0.76	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75
		Error spread over 6 months															
		0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	11%	12%	13%	14%	15%
Outcome Values	One %	0.28	0.25	0.19	0.16	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14
	Two %	0.53	0.48	0.38	0.30	0.27	0.27	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26
	Three %	0.71	0.67	0.54	0.43	0.38	0.36	0.36	0.36	0.36	0.36	0.36	0.36	0.35	0.36	0.36	0.36
	Four %	0.84	0.80	0.68	0.55	0.47	0.43	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42	0.42
	Five %	0.92	0.89	0.80	0.66	0.55	0.49	0.47	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46	0.46
	Six %	0.97	0.95	0.88	0.76	0.63	0.54	0.49	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48
	Seven %	0.99	0.98	0.94	0.84	0.71	0.59	0.52	0.50	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49
	Eight %	0.99	0.99	0.97	0.91	0.80	0.66	0.56	0.51	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
	Nine %	1.00	1.00	0.99	0.95	0.87	0.74	0.61	0.54	0.51	0.50	0.50	0.50	0.50	0.50	0.50	0.50
	Ten %	1.00	1.00	1.00	0.98	0.93	0.83	0.69	0.58	0.52	0.50	0.50	0.50	0.50	0.50	0.50	0.50
		Error spread over all months															
		0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	11%	12%	13%	14%	15%
Outcome Values	One %	0.28	0.27	0.22	0.16	0.10	0.06	0.03	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Two %	0.53	0.51	0.43	0.33	0.22	0.14	0.08	0.04	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	Three %	0.72	0.70	0.61	0.50	0.36	0.24	0.15	0.08	0.04	0.02	0.01	0.00	0.00	0.00	0.00	0.00
	Four %	0.85	0.83	0.77	0.65	0.52	0.38	0.25	0.16	0.09	0.05	0.02	0.01	0.00	0.00	0.00	0.00
	Five %	0.93	0.92	0.87	0.79	0.67	0.54	0.40	0.27	0.17	0.09	0.05	0.02	0.01	0.00	0.00	0.00
	Six %	0.97	0.97	0.94	0.88	0.80	0.69	0.55	0.41	0.28	0.18	0.10	0.05	0.02	0.01	0.00	0.00
	Seven %	0.99	0.99	0.97	0.95	0.89	0.82	0.70	0.57	0.43	0.30	0.19	0.11	0.06	0.03	0.01	0.00
	Eight %	1.00	0.99	0.99	0.98	0.95	0.90	0.83	0.72	0.60	0.46	0.32	0.21	0.12	0.07	0.03	0.01
	Nine %	1.00	1.00	1.00	0.99	0.98	0.96	0.91	0.84	0.75	0.62	0.48	0.35	0.23	0.14	0.08	0.04
	Ten %	1.00	1.00	1.00	1.00	0.99	0.98	0.96	0.92	0.86	0.77	0.65	0.52	0.38	0.26	0.16	0.09

Table 8.15: Experiment Three with the investigation rule 2 and data simulated to be like PGO3

When the error is zero, only if outcome values are greater than 4% will the level of false-negative signal be less than 0.1. When error is seeded into the audited twelve month period, then as with the earlier experiments, the performance of the AP varied with the pattern of that error. If the error was in relatively large amounts and, consequently was seeded into only a few months, then there were false-negative signals if $C \leq 4\%$. These can be seen in Table 8.15. There was a similar result if the error was spread over six months ($C \leq 3\%$). When the total error was dispersed across the year then the investigation rule had false-negative and false-positive signals while outcome-value and error were close: that is $|C - E| \leq 3\%$. If outcome-value was less than the error ($C < E$) then the signals were false-positive and, otherwise, the signals were false-negative.

It is clear from comparing Table 8.15 with Table 8.14 that this type of investigation-rule was much more effective when used with disaggregated data. The data was from the same source in each experiment. The result supports the expected consequence of disaggregating the data (the staff numbers in this experiment). The result is consistent with the guidance on the use of disaggregated information given to auditors in paragraph 13 of SAS 410. Using such data can produce an investigation-rule that better models the situation being audited, with the end result that there should be less false signals.

8.8 Experiment 4: The use of Bivariate Regression with data generated to be similar to that used by Knechel

This experiment considers the case when the data is simulated to be similar to that used in the work of Knechel (1988a) which was described in Chapter 7. The details of the data simulation are described in Section 7.7. In terms of the explanatory variables it is similar to experiments 1 and 2: that is there is one explanatory one variable. The purpose of these simulations was so that comparisons could be made to Knechel's work. The results of this experiment are shown in Table 8.16.

The first observation is that when the error is zero, only if outcome values are greater than 5% will the level of false-negative signal be less than 0.1. Considering the case when the total error is less than 5% of the value being audited, it can be seen that the results of this AP are again dependent on the distribution of the error. If the error is spread over all

months then, in terms of the combinations of error ($E \leq 5\%$) and all outcome-values ($C \leq 10\%$), there are a high percentage of false-positive signals (68% of the combinations of error and outcome-value). If the error is spread over 6 months there are false-negative signals while $C < 3\%$ and there are false-positive signals for other outcome-values while $1\% < E < 5\%$. For the other patterns of error there are false-negative signals while $C < 6\%$.

Outcome Values		Error spread over 1 month															
		0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	11%	12%	13%	14%	15%
	One %	0.24	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21
	Two %	0.45	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41	0.41
	Three %	0.63	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.58
	Four %	0.77	0.70	0.70	0.71	0.71	0.71	0.71	0.71	0.71	0.70	0.70	0.70	0.71	0.70	0.70	0.70
	Five %	0.86	0.79	0.79	0.79	0.79	0.79	0.79	0.79	0.79	0.79	0.79	0.79	0.79	0.79	0.79	0.79
	Six %	0.92	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85
	Seven %	0.96	0.89	0.88	0.88	0.88	0.89	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88
	Eight %	0.98	0.91	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90	0.90
	Nine %	0.99	0.93	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91
	Ten %	1.00	0.94	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91
		Error spread over 3 months															
		0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	11%	12%	13%	14%	15%
	One %	0.24	0.20	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17
	Two %	0.45	0.39	0.34	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.34	0.34	0.33
	Three %	0.63	0.56	0.48	0.47	0.47	0.47	0.47	0.46	0.46	0.47	0.46	0.47	0.46	0.47	0.47	0.46
	Four %	0.77	0.70	0.60	0.57	0.57	0.57	0.57	0.56	0.56	0.57	0.57	0.57	0.57	0.57	0.57	0.56
	Five %	0.86	0.80	0.70	0.65	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.64	0.65	0.64
	Six %	0.92	0.88	0.77	0.70	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69	0.69
	Seven %	0.96	0.93	0.83	0.74	0.72	0.72	0.72	0.72	0.72	0.72	0.72	0.72	0.72	0.72	0.72	0.72
	Eight %	0.98	0.96	0.88	0.78	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74	0.74
	Nine %	0.99	0.98	0.92	0.81	0.75	0.74	0.75	0.75	0.75	0.75	0.74	0.75	0.74	0.75	0.74	0.74
	Ten %	1.00	0.99	0.95	0.85	0.77	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.75
		Error spread over 6 months															
		0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	11%	12%	13%	14%	15%
	One %	0.24	0.21	0.17	0.14	0.12	0.12	0.11	0.11	0.11	0.11	0.12	0.12	0.11	0.12	0.12	0.24
	Two %	0.45	0.41	0.35	0.28	0.24	0.23	0.22	0.22	0.21	0.22	0.22	0.22	0.22	0.22	0.22	0.45
	Three %	0.63	0.59	0.50	0.40	0.34	0.32	0.32	0.30	0.31	0.31	0.31	0.31	0.31	0.31	0.31	0.63
	Four %	0.77	0.73	0.64	0.53	0.44	0.40	0.39	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.38	0.77
	Five %	0.86	0.83	0.76	0.64	0.54	0.47	0.44	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.43	0.86
	Six %	0.92	0.91	0.84	0.74	0.63	0.54	0.49	0.47	0.47	0.46	0.46	0.46	0.46	0.46	0.46	0.92
	Seven %	0.96	0.95	0.91	0.82	0.70	0.60	0.53	0.50	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.96
	Eight %	0.98	0.98	0.95	0.88	0.78	0.68	0.58	0.52	0.50	0.49	0.49	0.49	0.49	0.49	0.49	0.98
	Nine %	0.99	0.99	0.97	0.93	0.85	0.74	0.63	0.56	0.52	0.50	0.50	0.50	0.50	0.50	0.50	0.99
	Ten %	1.00	1.00	0.99	0.96	0.91	0.81	0.70	0.60	0.54	0.51	0.50	0.50	0.50	0.50	0.50	1.00
		Error spread over all months															
		0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	11%	12%	13%	14%	15%
	One %	0.24	0.23	0.20	0.17	0.12	0.08	0.05	0.03	0.02	0.01	0.01	0.00	0.00	0.00	0.00	0.00
	Two %	0.45	0.44	0.39	0.33	0.26	0.18	0.12	0.07	0.04	0.02	0.01	0.01	0.00	0.00	0.00	0.00
	Three %	0.63	0.62	0.56	0.48	0.38	0.29	0.20	0.13	0.08	0.05	0.03	0.01	0.01	0.00	0.00	0.00
	Four %	0.77	0.75	0.71	0.63	0.52	0.41	0.31	0.21	0.14	0.09	0.05	0.03	0.01	0.01	0.00	0.00
	Five %	0.86	0.86	0.81	0.75	0.66	0.54	0.43	0.32	0.23	0.14	0.09	0.05	0.03	0.02	0.01	0.00
	Six %	0.92	0.92	0.90	0.84	0.77	0.68	0.56	0.44	0.34	0.24	0.15	0.10	0.06	0.03	0.02	0.01
	Seven %	0.96	0.96	0.94	0.91	0.86	0.78	0.69	0.58	0.46	0.35	0.25	0.16	0.10	0.06	0.03	0.02
	Eight %	0.98	0.98	0.97	0.95	0.92	0.87	0.79	0.70	0.60	0.48	0.37	0.27	0.17	0.11	0.07	0.04
	Nine %	0.99	0.99	0.99	0.98	0.96	0.93	0.88	0.81	0.72	0.62	0.50	0.39	0.29	0.19	0.12	0.08
	Ten %	1.00	1.00	1.00	0.99	0.99	0.97	0.94	0.89	0.82	0.75	0.64	0.53	0.41	0.31	0.21	0.14

Table 8.16: Experiment Four with the investigation rule 1 and data simulated to be similar to that used by Knechel (1988a)

Comparing these results with those of Knechel (1988a), we see that the experiment described here clearly highlights an issue disguised by Knechel's measurement of the effectiveness of the AP, which used the reduced sample sizes of subsequent substantive tests. That reduction was calculated using the Audit Risk Model. Knechel did not identify the high levels of false-positive signals: in his experiments there were no increases in sample sizes to signify that such false signals had been identified. The implications of this are discussed section 8.10.

8.9 Experiments 5 to 11: Results for Simple APs using Annual Outcome-rules

This section describes the experiments carried out to investigate whether simple APs, using data from the pay area, are effective when the criteria for effectiveness are false-negative and false-positive signals. Also the experiments investigate whether such APs will replicate the findings of Loebbecke and Steinbart (1987) when used with data from a different source to theirs, or will confirm the finding of Knechel (1988b). Loebbecke and Steinbart's work was described in Section 7.5 and that of Knechel in section 7.8.

The simple rules used in this sequence of experiments are shown in Table 8.17. The findings of experiments 5 to 11 are presented in four sections 8.9.1 to 8.9.4. Within each of these sections the investigation-rules are similar to each other.

Rule	Description of the investigation-rule	Results
	Comparison	**In Table**
1.3	Compare to last year	8.18
	Prediction using percentage change in previous years	
1.4	Year 4 = Year 3 × (1 + % change of years 3 and 2)	8.19
1.5	Year 4 = Year 3 × (Average of % change of years 3 & 2 and years 2 & 1)	8.20
	Prediction using average change in previous years	
1.6	Year 4 = Year 3 + (Year 2 – Year 1)	8.21
1.7	Year 4 = Year 3 + {(Year 3 – Year 2) + (Year 2– Year 1}/2	8.22
1.8	Year 4 = Year 3 + {2× (Year 3 – Year 2) + 1× (Year 2– Year 1}/3	8.23
	Prediction using average of previous years	
1.9	Year 4 = {(Year 1 + Year 2 + Year 3)/3}	8.24

Table 8.17: The Investigation Rules for experiments five to eleven

The data was simulated in the same way as the in experiments 1 and 2, and then summed to produce annual totals. As a result there were sets of four annual totals, where years one

to three represented the period prior to the audit, and year four was the value of the audited year. This was done for two reasons. Firstly to have data that was consistent with the experiments 1 and 2 and secondly because they are typical of those used by many auditors.

In each of the tables reporting the results (e.g. Table 8.18), the first half of the table is the experiment that used the data simulated to similar to that of the PGO Agency and the second half of the table is the data from the LR Agency.

8.9.1 Simple comparison type investigation rule (Table 8.18)

Investigation rule 1.3 was a very simple rule comparing the current year to the previous year. As can be seen this investigation-rule produced a large number of false-signals. While the error was less than the outcome-rule, there were high levels of false-negative signals for almost all values of the outcome-value (C). Considering the results while $E \leq 5\%$, there were false-negative signals for most combinations of error and outcome-rule. The results are shown in Table 8.18 below.

Outcome Values	PGO	Error Rates															
		0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	11%	12%	13%	14%	15%
	One %	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Two %	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Three %	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Four %	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Five %	0.06	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Six %	0.27	0.06	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Seven %	0.58	0.29	0.08	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Eight %	0.83	0.60	0.32	0.09	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Nine %	0.96	0.84	0.63	0.36	0.12	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Ten %	0.99	0.96	0.86	0.66	0.40	0.14	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Outcome Values	LR	Error Rates															
		0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	11%	12%	13%	14%	15%
	One %	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Two %	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Three %	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Four %	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Five %	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Six %	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Seven %	0.12	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Eight %	0.30	0.13	0.03	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Nine %	0.52	0.32	0.14	0.04	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Ten %	0.74	0.55	0.35	0.16	0.04	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table 8.18: Investigation Rule 1.3: comparison.

8.9.2 Simple percentage change type investigation-rule (Tables 8.19 & 8.20)

The next experiments applied 'simple percentage change' investigation-rules (rules 1.4 and 1.5 of Table 8.17). The results are shown in Table 8.19 and Table 8.20 respectively. The rules are given at the foot of the relevant table.

	PGO	Error Rates															
		0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	11%	12%	13%	14%	15%
Outcome Values	One %	0.35	0.38	0.30	0.19	0.11	0.04	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Two %	0.65	0.65	0.58	0.42	0.24	0.13	0.05	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Three %	0.83	0.84	0.76	0.62	0.44	0.26	0.14	0.06	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	Four %	0.94	0.94	0.89	0.78	0.64	0.46	0.27	0.15	0.06	0.02	0.01	0.00	0.00	0.00	0.00	0.00
	Five %	0.98	0.98	0.96	0.90	0.79	0.66	0.48	0.29	0.16	0.07	0.03	0.01	0.00	0.00	0.00	0.00
	Six %	0.99	0.99	0.99	0.96	0.91	0.80	0.68	0.50	0.31	0.18	0.09	0.03	0.01	0.00	0.00	0.00
	Seven %	1.00	1.00	1.00	0.99	0.97	0.92	0.82	0.70	0.52	0.34	0.20	0.10	0.04	0.01	0.01	0.00
	Eight %	1.00	1.00	1.00	1.00	0.99	0.97	0.93	0.84	0.71	0.56	0.38	0.22	0.12	0.05	0.02	0.01
	Nine %	1.00	1.00	1.00	1.00	1.00	0.99	0.97	0.94	0.85	0.74	0.59	0.41	0.25	0.14	0.06	0.02
	Ten %	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.98	0.95	0.88	0.77	0.63	0.46	0.28	0.16	0.08
	LR	Error Rates															
		0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	11%	12%	13%	14%	15%
Outcome Values	One %	0.25	0.28	0.25	0.21	0.16	0.11	0.07	0.04	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	Two %	0.49	0.51	0.49	0.41	0.32	0.23	0.15	0.09	0.05	0.02	0.01	0.01	0.00	0.00	0.00	0.00
	Three %	0.67	0.70	0.67	0.60	0.49	0.36	0.26	0.17	0.10	0.06	0.03	0.01	0.01	0.00	0.00	0.00
	Four %	0.81	0.83	0.80	0.74	0.65	0.52	0.39	0.27	0.18	0.11	0.06	0.03	0.02	0.01	0.00	0.00
	Five %	0.89	0.91	0.90	0.85	0.77	0.66	0.54	0.40	0.29	0.19	0.12	0.07	0.04	0.02	0.01	0.01
	Six %	0.94	0.96	0.95	0.92	0.86	0.78	0.68	0.56	0.42	0.31	0.20	0.13	0.08	0.04	0.02	0.01
	Seven %	0.98	0.98	0.98	0.96	0.93	0.87	0.79	0.70	0.57	0.44	0.32	0.22	0.14	0.09	0.05	0.02
	Eight %	0.99	0.99	0.99	0.98	0.97	0.93	0.88	0.81	0.71	0.60	0.46	0.35	0.24	0.16	0.10	0.06
	Nine %	0.99	1.00	1.00	0.99	0.99	0.97	0.94	0.89	0.82	0.73	0.62	0.49	0.37	0.26	0.18	0.11
	Ten %	1.00	1.00	1.00	1.00	1.00	0.99	0.98	0.95	0.90	0.84	0.75	0.65	0.53	0.39	0.29	0.20

Table 8.19: Investigation rule 1.4: Year 4 = Year 3 × (1 + % change of years 3 and 2)

	PGO	Error Rates															
		0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	11%	12%	13%	14%	15%
Outcome Values	One %	0.34	0.37	0.32	0.24	0.15	0.06	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Two %	0.63	0.66	0.61	0.47	0.30	0.17	0.07	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Three %	0.81	0.87	0.81	0.68	0.49	0.32	0.18	0.07	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Four %	0.93	0.96	0.92	0.83	0.69	0.51	0.34	0.19	0.08	0.02	0.00	0.00	0.00	0.00	0.00	0.00
	Five %	0.98	0.98	0.97	0.93	0.84	0.71	0.53	0.36	0.20	0.09	0.03	0.00	0.00	0.00	0.00	0.00
	Six %	1.00	1.00	0.99	0.97	0.94	0.85	0.73	0.55	0.38	0.22	0.11	0.04	0.01	0.00	0.00	0.00
	Seven %	1.00	1.00	1.00	0.99	0.97	0.94	0.87	0.75	0.58	0.41	0.24	0.13	0.05	0.01	0.00	0.00
	Eight %	1.00	1.00	1.00	1.00	0.99	0.98	0.95	0.88	0.77	0.61	0.44	0.27	0.15	0.06	0.02	0.00
	Nine %	1.00	1.00	1.00	1.00	1.00	0.99	0.98	0.95	0.89	0.79	0.65	0.47	0.31	0.18	0.08	0.02
	Ten %	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.98	0.96	0.91	0.82	0.68	0.51	0.35	0.20	0.10
	LR	Error Rates															
		0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	11%	12%	13%	14%	15%
Outcome Values	One %	0.23	0.29	0.29	0.26	0.21	0.15	0.08	0.04	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	Two %	0.46	0.51	0.54	0.50	0.41	0.30	0.19	0.11	0.06	0.02	0.01	0.00	0.00	0.00	0.00	0.00
	Three %	0.64	0.71	0.72	0.68	0.58	0.46	0.33	0.21	0.12	0.06	0.03	0.01	0.00	0.00	0.00	0.00
	Four %	0.78	0.85	0.85	0.81	0.73	0.61	0.48	0.34	0.22	0.13	0.07	0.03	0.01	0.01	0.00	0.00
	Five %	0.89	0.93	0.93	0.89	0.83	0.74	0.63	0.50	0.35	0.24	0.14	0.08	0.03	0.01	0.01	0.00
	Six %	0.95	0.97	0.97	0.95	0.91	0.84	0.75	0.64	0.51	0.37	0.26	0.16	0.08	0.04	0.02	0.01
	Seven %	0.98	0.99	0.99	0.98	0.95	0.92	0.85	0.77	0.66	0.53	0.39	0.28	0.18	0.09	0.05	0.02
	Eight %	0.99	1.00	0.99	0.99	0.98	0.96	0.92	0.87	0.78	0.68	0.56	0.42	0.30	0.19	0.11	0.06
	Nine %	1.00	1.00	1.00	1.00	0.99	0.98	0.96	0.93	0.88	0.80	0.71	0.59	0.45	0.33	0.22	0.13
	Ten %	1.00	1.00	1.00	1.00	1.00	0.99	0.98	0.97	0.94	0.89	0.81	0.73	0.61	0.49	0.35	0.24

Table 8.20: Investigation rule 1.5: Year 4 = Year 3 × (Average of % change of years 3 and 2 and years 2 and 1)

It can be seen that the results are very similar. While outcome-value and error close, there were false-negative signals if C > E and false-positive signals if C < E. In this case close meant that $|C - E| \leq 3\%$ for PGO data and $|C - E| \leq 4\%$ for LR data.

The false-positive signals are particularly worrying since they will cause the auditor to wrongly conclude that the AP has supported their prior judgement that, for example, the business systems would prevent or detect significant errors: in these cases significant means greater than the outcome-value used.

8.9.3 Simple average change type investigation rule (Tables 8.21, 8.22 & 8.23)

The results of applying the average change investigation-rules (rules 1.6, 1.7 and 1.8 of Table 8.17) are shown in Tables 8.21, 8.22 and 8.23 respectively.

	PGO	Error Rates															
		0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	11%	12%	13%	14%	15%
Outcome Values	One %	0.36	0.31	0.22	0.16	0.08	0.03	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Two %	0.63	0.58	0.47	0.31	0.19	0.09	0.03	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Three %	0.83	0.78	0.67	0.50	0.32	0.20	0.10	0.04	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	Four %	0.94	0.91	0.81	0.68	0.52	0.34	0.21	0.11	0.04	0.02	0.01	0.00	0.00	0.00	0.00	0.00
	Five %	0.98	0.96	0.92	0.83	0.70	0.53	0.36	0.23	0.12	0.05	0.02	0.01	0.00	0.00	0.00	0.00
	Six %	0.99	0.99	0.97	0.93	0.84	0.71	0.55	0.38	0.24	0.13	0.05	0.02	0.01	0.00	0.00	0.00
	Seven %	1.00	1.00	0.99	0.97	0.93	0.85	0.73	0.58	0.41	0.26	0.14	0.07	0.03	0.01	0.00	0.00
	Eight %	1.00	1.00	1.00	0.99	0.97	0.94	0.87	0.75	0.61	0.44	0.28	0.16	0.09	0.03	0.01	0.01
	Nine %	1.00	1.00	1.00	1.00	0.99	0.98	0.95	0.88	0.77	0.64	0.47	0.31	0.19	0.10	0.04	0.02
	Ten %	1.00	1.00	1.00	1.00	1.00	0.99	0.98	0.95	0.90	0.79	0.66	0.51	0.34	0.22	0.12	0.05
	LR	Error Rates															
		0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	11%	12%	13%	14%	15%
Outcome Values	One %	0.24	0.25	0.23	0.16	0.11	0.07	0.04	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Two %	0.49	0.48	0.41	0.33	0.23	0.15	0.09	0.05	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	Three %	0.68	0.65	0.59	0.48	0.38	0.25	0.16	0.10	0.05	0.03	0.01	0.00	0.00	0.00	0.00	0.00
	Four %	0.82	0.79	0.72	0.63	0.51	0.40	0.27	0.17	0.10	0.06	0.03	0.01	0.00	0.00	0.00	0.00
	Five %	0.90	0.89	0.83	0.75	0.64	0.53	0.41	0.29	0.18	0.11	0.06	0.03	0.01	0.01	0.00	0.00
	Six %	0.95	0.94	0.91	0.85	0.75	0.65	0.54	0.42	0.31	0.20	0.12	0.07	0.04	0.02	0.01	0.00
	Seven %	0.98	0.97	0.95	0.92	0.86	0.76	0.67	0.56	0.44	0.33	0.21	0.13	0.08	0.04	0.02	0.01
	Eight %	0.99	0.99	0.98	0.95	0.92	0.87	0.78	0.68	0.57	0.46	0.35	0.23	0.14	0.09	0.05	0.02
	Nine %	1.00	1.00	0.99	0.98	0.96	0.93	0.88	0.79	0.70	0.60	0.48	0.37	0.25	0.16	0.10	0.06
	Ten %	1.00	1.00	1.00	0.99	0.98	0.96	0.94	0.89	0.81	0.72	0.62	0.50	0.40	0.28	0.18	0.12

Table 8.21: Investigation rule 1.6: Year 4 = Year 3 + (Year 2 – Year 1)

For the average change type of investigation rule set out in Figure 8.17 it can be seen that, as in Section 8.9.2, while outcome-value and error close, there were false-negative signals while C > E and false-positive signals while C < E. What constituted close varied with the outcome-rule and error. For example in Table 8.21 and 8.23, for PGO data close was never less than 2% and for LR data close was never less than 3%, while in Table 8.22 it was never less than 2% (for both PGO and LR data). It could be much higher, for example in

Table 8.22 $|C - E| \leq 6\%$ when outcome-value was in the range 4% to 10% and data was simulated to be like that of LR. Again for all these experiments the comment about the levels of false-positive signals applies.

Outcome Values	PGO	Error Rates															
		0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	11%	12%	13%	14%	15%
	One %	0.40	0.37	0.26	0.13	0.05	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Two %	0.72	0.66	0.50	0.31	0.15	0.05	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Three %	0.89	0.85	0.71	0.53	0.32	0.16	0.06	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Four %	0.97	0.94	0.86	0.72	0.54	0.34	0.17	0.07	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Five %	0.99	0.98	0.95	0.87	0.74	0.56	0.36	0.19	0.08	0.02	0.00	0.00	0.00	0.00	0.00	0.00
	Six %	1.00	0.99	0.98	0.95	0.88	0.75	0.58	0.38	0.20	0.09	0.02	0.01	0.00	0.00	0.00	0.00
	Seven %	1.00	1.00	1.00	0.98	0.95	0.89	0.77	0.60	0.41	0.22	0.11	0.03	0.01	0.00	0.00	0.00
	Eight %	1.00	1.00	1.00	1.00	0.98	0.96	0.90	0.79	0.62	0.45	0.26	0.12	0.04	0.01	0.00	0.00
	Nine %	1.00	1.00	1.00	1.00	1.00	0.98	0.96	0.91	0.82	0.66	0.49	0.29	0.15	0.06	0.02	0.00
	Ten %	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.97	0.93	0.84	0.70	0.53	0.34	0.18	0.08	0.02
Outcome Values	LR	Error Rates															
		0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	11%	12%	13%	14%	15%
	One %	0.31	0.30	0.24	0.16	0.09	0.05	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Two %	0.58	0.55	0.46	0.33	0.21	0.11	0.05	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Three %	0.78	0.74	0.64	0.51	0.36	0.22	0.12	0.06	0.03	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	Four %	0.89	0.87	0.78	0.67	0.53	0.37	0.23	0.12	0.06	0.03	0.01	0.00	0.00	0.00	0.00	0.00
	Five %	0.96	0.94	0.89	0.80	0.68	0.54	0.38	0.25	0.14	0.07	0.03	0.01	0.00	0.00	0.00	0.00
	Six %	0.98	0.97	0.94	0.90	0.81	0.69	0.55	0.40	0.26	0.15	0.08	0.04	0.01	0.00	0.00	0.00
	Seven %	1.00	0.99	0.98	0.95	0.90	0.82	0.71	0.57	0.43	0.28	0.17	0.09	0.04	0.02	0.01	0.00
	Eight %	1.00	1.00	0.99	0.98	0.95	0.91	0.83	0.73	0.59	0.45	0.31	0.19	0.10	0.05	0.02	0.01
	Nine %	1.00	1.00	1.00	0.99	0.98	0.96	0.92	0.85	0.75	0.62	0.48	0.34	0.21	0.12	0.06	0.03
	Ten %	1.00	1.00	1.00	1.00	0.99	0.98	0.96	0.92	0.86	0.77	0.65	0.52	0.37	0.24	0.14	0.07

Table 8.22: Investigation rule 1.7: Year 4 = Year 3 + {(Year 3 – Year 2) + (Year 2– Year 1}/2

Outcome Values	PGO	Error Rates															
		0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	11%	12%	13%	14%	15%
	One %	0.42	0.38	0.25	0.13	0.05	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Two %	0.73	0.67	0.51	0.31	0.15	0.06	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Three %	0.90	0.85	0.72	0.53	0.32	0.16	0.07	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Four %	0.97	0.95	0.87	0.74	0.54	0.34	0.17	0.07	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	Five %	0.99	0.99	0.95	0.88	0.75	0.56	0.36	0.19	0.08	0.03	0.01	0.00	0.00	0.00	0.00	0.00
	Six %	1.00	1.00	0.99	0.96	0.89	0.77	0.59	0.39	0.21	0.09	0.04	0.01	0.00	0.00	0.00	0.00
	Seven %	1.00	1.00	1.00	0.99	0.96	0.90	0.79	0.61	0.42	0.24	0.11	0.04	0.01	0.00	0.00	0.00
	Eight %	1.00	1.00	1.00	1.00	0.99	0.97	0.91	0.80	0.64	0.45	0.26	0.12	0.05	0.02	0.00	0.00
	Nine %	1.00	1.00	1.00	1.00	1.00	0.99	0.97	0.92	0.82	0.67	0.49	0.30	0.15	0.07	0.02	0.01
	Ten %	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.97	0.93	0.85	0.71	0.53	0.34	0.18	0.08	0.03
Outcome Values	LR	Error Rates															
		0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	11%	12%	13%	14%	15%
	One %	0.32	0.30	0.25	0.18	0.10	0.05	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Two %	0.58	0.56	0.48	0.35	0.22	0.12	0.05	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Three %	0.77	0.76	0.67	0.53	0.37	0.23	0.13	0.06	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	Four %	0.91	0.87	0.81	0.68	0.53	0.39	0.24	0.14	0.07	0.03	0.01	0.00	0.00	0.00	0.00	0.00
	Five %	0.96	0.95	0.89	0.81	0.70	0.55	0.40	0.26	0.15	0.07	0.03	0.01	0.00	0.00	0.00	0.00
	Six %	0.99	0.98	0.95	0.89	0.82	0.71	0.56	0.42	0.28	0.17	0.09	0.03	0.01	0.00	0.00	0.00
	Seven %	1.00	0.99	0.98	0.96	0.90	0.83	0.73	0.59	0.44	0.29	0.18	0.10	0.04	0.02	0.01	0.00
	Eight %	1.00	1.00	0.99	0.98	0.96	0.91	0.84	0.74	0.61	0.47	0.32	0.20	0.11	0.05	0.02	0.01
	Nine %	1.00	1.00	1.00	0.99	0.98	0.96	0.92	0.86	0.77	0.64	0.50	0.35	0.22	0.13	0.06	0.03
	Ten %	1.00	1.00	1.00	1.00	0.99	0.98	0.97	0.93	0.87	0.79	0.67	0.53	0.39	0.25	0.15	0.08

Table 8.23: Investigation rule 1.8: Year 4 = Year 3 + {2× (Year 3 – Year 2) + 1×(Year 2– Year 1}/3

8.9.4 Simple average of previous years type investigation rule (Table 8.24)

The results of applying the investigation-rule, 'average of previous years' (rule 1.9 of Table 8.17) are shown in Table 8.24 below. Inspection of the table reveals that while E < 5% and C < 15% there are high levels of false-negative signals. It should be noted that this investigation-rule was evaluated for outcome-values much greater than the other experiments. For outcome-values less than 10% there were high levels of false signals, consequently the range of outcome-values was increased to investigate at what level the frequency of false signals became low (< 0.1 – light blue in the tables). An investigation-rule such as this is a problem for the audit process if auditors react to the signals by undertaking further work that would prove to be nugatory.

Outcome Values	PGO	Error Rates															
		0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	11%	12%	13%	14%	15%
	1%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	5%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	10%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	15%	0.89	0.68	0.39	0.13	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	20%	1.00	1.00	1.00	1.00	1.00	0.98	0.91	0.76	0.49	0.22	0.07	0.01	0.00	0.00	0.00	0.00
	21%	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.94	0.82	0.59	0.32	0.11	0.03	0.00	0.00	0.00
	22%	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.97	0.88	0.70	0.42	0.17	0.05	0.01	0.00
	23%	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.98	0.93	0.78	0.55	0.28	0.09	0.02
	24%	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.96	0.87	0.68	0.40	0.17
	25%	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.98	0.92	0.78	0.54
Outcome Values	LR	Error Rates															
		0%	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	11%	12%	13%	14%	15%
	1%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	5%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	10%	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	15%	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	20%	0.91	0.80	0.61	0.39	0.21	0.09	0.03	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	21%	0.98	0.93	0.82	0.67	0.45	0.25	0.12	0.04	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	22%	1.00	0.99	0.95	0.86	0.73	0.53	0.32	0.16	0.07	0.02	0.01	0.00	0.00	0.00	0.00	0.00
	23%	1.00	1.00	0.99	0.97	0.90	0.79	0.61	0.39	0.22	0.10	0.03	0.01	0.00	0.00	0.00	0.00
	24%	1.00	1.00	1.00	0.99	0.98	0.93	0.83	0.68	0.47	0.28	0.14	0.06	0.02	0.00	0.00	0.00
	25%	1.00	1.00	1.00	1.00	1.00	0.99	0.96	0.88	0.76	0.57	0.36	0.20	0.09	0.03	0.01	0.00

Table 8.24: Investigation rule 1.9: Year 4 = {(Year 1 + Year 2 + Year 3)/3}

8.10 Overall Summary and Conclusions for Chapter 8

This chapter has reported some new experiments that explored, for different combinations of seeded error ($0\% \leq E \leq 15\%$) and outcome-values ($1\% \leq C \leq 10\%$), the performance of a number of investigation-rules. Both the seeded error and the outcome-values were a percentage of the value being audited (Y_B). For these experiments a new criterion was used

to measure the effectiveness of the investigation rules. It was the levels of false-negative and false-positive signals.

8.10.1 The investigation-rules and outcome-rules

There were two classes of investigation-rules used in the experiments. The first was of a regression type, similar to those used by some auditors (e.g. the STAR package of Deloitte and Touche). Despite the fact the surveys (described in Chapters 5 and 6) reported that such rules are not widely used, it was decided to use regression type investigation-rules since auditors do use them, albeit infrequently, and because they are representative of the more complex type of AP applied by auditors. Another reason was that such rules are similar to those used by Knechel (1988a) and Kaplan (1978) and it was necessary to test their findings. The other class of investigation-rules were simple rules that are typical of those used extensively by auditors. They were analogous to those used by Loebbecke and Steinbart (1987) and Knechel (1988b). All the investigation-rules are summarised in Table 8.2 on page 188.

In Section 8.2.1 (page 186), two criteria were discussed for measuring the closeness of the estimated-value to the book-value. They were 'auditing consistency' and 'statistical consistency'. In these experiments, it was decided not to use the statistical consistency measure. There were two reasons for this decision. Firstly, while some of the APs allowed quantified estimates of precision, others did not, and the intention was to compare all the APs on an even footing. Secondly, auditors do not normally consider statistical consistency, and it was decided to emulate that situation.

The experiments therefore concentrated on auditing consistency which was defined in Section 8.2.1 to be that $|\hat{Y}_T - Y_B|$ is less than some significant amount such as the outcome-value. Although by their own admission, for example in the survey reported in Chapter 6, many auditors do not have a pre-defined value for an auditing consistency rule, they do make a comparison between $\hat{Y}_T$ and Y_B. So the use of such a rule in the experiments is consistent with the audit practice, both the formal and informal uses of such a rule.

Despite the decision not to use statistical consistency, and hence not use the precision of the estimates, there is an important caveat to the results of the experiments. It is that for many of the investigation-rules used in the experiments the precision would be greater than the outcome-values applied by the auditor! This may well be a consequence of the type of rule used by auditors and the short runs of data. It raises the dilemma that it is not useful to the auditor to use a procedure that is, potentially, less precise than the outcome-value they purport to use.

8.10.2 The data used as a basis for the simulations

The investigation-rules were applied to three different sets of monthly data. For each of the three populations, 2,500 sets of data were simulated. The first two data populations were simulated to have the characteristics of monthly pay data from two Government Agencies. These data were chosen to see how the investigation-rules would perform on data that was from a well-controlled, stable, area of many organisations: the pay of staff. If an investigation-rule is likely to perform well it should do so with such data. If it does not, then that will be a concern, since it will imply that APs using similar investigation-rules could be unreliable. A rule that performs badly with good quality data is unlikely to perform better with poor data.

The third set of data was simulated to be similar to that created by Knechel (1988a). The data was increasing, in a structured way, with a random error that was distributed as if it came from a normal distribution: the details that underlie its generation are described in Section 7.7. The reason for using this data was to assess how regression based investigation-rules would perform when the criterion for effectiveness was measured directly using false signals, rather than by the indirect measurement used by Knechel: the reduction of other substantive tests.

8.10.3 The results: monthly data and regression type investigation-rules

For regression type investigation-rules, used in conjunction with monthly outcome-rules when no error was seeded into the audited populations, the performance varied between investigation-rules and data sources. Generally there were high levels of false-negative

signals (coloured dark blue in the tables) for outcome-value less than 1% to 4%: the lower figure being achieved with disaggregated data. The equivalent figures for low frequencies of false-negative signals (coloured light blue in the tables) were between 4% and 10% respectively. These results apply to all the data sources used in the experiments described in this chapter.

If the data simulated to be analogous to the two real populations, and when the error was seeded into the data and spread over 1, 3 or 6 months, there were false signals for many combinations of error and outcome-value. When the error was spread over all months, then the signals were false-negative while the outcome-value was greater than the seeded error, and false-positive while outcome-value was less than the seeded error.

When the regression type of investigation-rule was used on data that was similar to Knechel's and the error was spread over 1, 3 or 6 months, the investigation-rules performed better than they did when applied to the Agency data. They had lower levels of false-negative signals. However for the Knechel type data, when the error was spread over 12 months, there were many false-positive signals.

It is possible to make a comparison between the results of the experiments reported in sections 8.5 to 8.8 and those of Knechel (1988a). This is because the same type of data and investigation-rules were used. From an inspection of the results of the experiments in sections 8.5 to 8.8 it can be seen that, for many combinations of error and outcome-value, the investigation-rules perform quite badly when effectiveness is measured directly by the level of false signals. This result does not agree with the findings of Knechel (1988a), whose measure of effectiveness was the potential reduction in sample sizes.

As was observed in Chapter 7, Kaplan (1978) found that his prediction did not come within 10% of the book-value being tested. That finding is consistent with the experiments reported in this chapter. Thus it is likely that Knechel's use of the Audit Risk Model disguised the observation of false signals, and therefore his finding that APs were effective may well be flawed. This is particularly significant for cases where the seeded error was greater than the outcome-value, and the error and outcome-value were close. Then there were false-positive signals, which could cause the wrong audit decision to be made.

The auditor would be unaware of high false-positive signals found by the experiments. As was seen in the surveys, such signals might persuade them to reduce their other tests and, as a consequence, lower the probability of detecting the presence of error. Alternatively the auditor might have already reduced their substantive tests, because they judged the systems to be good, and the use of the APs represents the only such test. Its failure to detect potential significant error could cause the wrong audit decision. Such conclusions imply that this type of APs might not be an effective procedure.

There are, however, instances where the results of the experiments agreed with those of Knechel. For example, when the error is spread over one or three months and the outcome-values are set to be $\geq 5\%$, then there will be low levels of false signals and so the results of the experiments will be consistent with Knechel's conclusion about effectiveness. Another instance is that the false-positive signals in the experiments reported in this chapter were lower if the error occurred in large amounts: that is, the error was not spread evenly through the year. This concurs with Knechel's (1988a) observation that: '*all APs were most effective when the data simulated large nonrecurring errors rather than small, frequent errors that were material.*'

8.10.4 The results: annual data and the simple investigation-rules

The second group of experiments, reported in Section 8.9 (in pages 210 to 215) applied a range of simple investigation-rules to the data generated in a similar way to the first set of experiments, but in order to replicate annual data it was summed to produce four annual amounts. Sections 8.9.1 to 8.9.4 describe the results for the simple investigations-rules.

As before, when no error was seeded into the audited populations, the performance varied between different investigation-rules and/or data sources. There were high levels of false-negative signals (coloured dark blue in the tables) when the outcome-values were between 1% and 15%. The false-negative signals generally occurred when the outcome-values were at the lower end of this interval. The equivalent figures for low frequencies of false-negative signals (coloured light blue in the tables) were between 3% and 15% respectively.

When error was seeded into the audited year, there were differing levels of false signals for all the investigation-rules. They occurred while the outcome-value (C) and error (E) were close. They were false-negative while C > E, and false-positive while C < E. For example, for investigation-rule 1.5 (Table 8.20 on page 212) 'close' was, at best, $|C - E| \leq 4\%$ for one set of agency data. The equivalent figure for the data from the other agency was, at best, $|C - E| \leq 3\%$. These values were typical of the results from the other experiments in this chapter. They are also consistent with the work of Loebbecke and Steinbart (1987) who found high frequencies of Type I and II error, the equivalent of the false-negative and false-positive signals used in this thesis. Thus an important finding is that the simple investigation-rules potentially have high occurrences of false signals when the outcome-value and error are close. Again a worrying feature is the levels of false-positive signals.

8.10.5 Concluding remarks

One of the contributions of this chapter is that it provides evidence about the levels of false signals. In most of the experiments, while error rates were low, there were varying degrees of false-negative signals. This was true for both simple and regression based investigation-rules and for monthly and annual outcome-rules. In particular, if there was no error, then for outcome-values of $\leq 4\%$ there were false-negative signals. For some of the investigation-rules there were false-negative signals when the outcome-values were in excess of 4%. These results imply either nugatory work for the auditor, or that outcome-values are set at greater than 4% of the value being audited by the AP. An implication is that auditors use high outcome-values. That deduction is supported by the survey done for this thesis where the auditors, when questioned, did not report nugatory work that would be the result of high frequencies of false-negative signals. The supposition that high outcome-values are used is also consistent with a conclusion of Kaplan (1978) who found that using investigation-rules based on regression never got closer than within 10% of the book value. So an implication is that auditors use high outcome-values, high that is relative to the particular audit test.

This raises a question about the auditor's use of high outcome-rules as a yardstick in the audit process. It is, how accurate is the audit and what is it achieving? This may well be

part of the expectation-gap and a question is whether the audit is conducted to the high standards that the users expect.

Another result was that when error was high, relative to the outcome-value, some of the investigation-rules produced false-positive signals. It was said in some of the research reported in Chapter 7 that certain APs, when used in conjunction with monthly outcome-rules, were not good at detecting error if the error occurred in small amounts, even if in total that error might be material. That situation is replicated in the experiments described in this chapter. For example when the error is spread over twelve months and a monthly outcome-rule was used, then the experiments identified high frequencies of false-positive signals. Such false-positive signals also occurred when simple APs were used with annual outcome-rules. These occurred when the error and outcome-value were close and, in addition, the error was greater than the outcome-value.

The finding about the level of false-positive signals is particularly worrying if the auditor is using APs as the sole source of substantive evidence. Then, those procedures are unlikely to detect that something is wrong with the auditor's prior judgement, or at least to obtain a signal to that effect. This situation could arise when the auditor has decided to place reliance of the business systems and to do a minimal substantive test – the AP. The experiments demonstrated that when the error is a little larger than the outcome-value, that the AP may provide false-positive signals and in these circumstances, the AP does not signal the potential error. An exception might be when the error is substantially greater than the outcome value.

Taken together all these results imply that APs, because they have false-negative and false-positive signals for many combinations of error and outcome-value, are not the effective tools that auditors perceived them to be. The investigations reported in this thesis imply that in certain circumstances, APs are ineffective, rather than cost effective.

Chapter 9: Overall Summary and Conclusions

9.1 Introduction

The purpose of an audit is to provide assurance about the 'truth and fairness' of a set of financial statements to the users of those statements. To do that, auditors collect a balance of evidence and information from many different sources. The Auditing Standards require that: *'the auditor obtains with reasonable assurance, sufficient reliable evidence that the financial statements are free of material error'* (SAS100). Substantive tests provide one form of evidence, and this thesis has been concerned primarily with one source of such evidence: APs, and in particular, whether APs can provide sufficient, reliable substantive evidence.

9.2 The requirement for substantive tests

The objectives of a substantive test are described in paragraph 106 of the Audit Guidance for the audit of UK Banks, which was issued by the Auditing Practices Committee (APC), the predecessor to the APB. It is quite clear and says: *'The objective of a substantive test is to gather evidence as to the completeness, accuracy and validity of the information contained in the financial statements'* (APC – AG 307 (1989)). Since then Statements of Auditing Standards have been issued, where the substantive tests are defined as: – '*Tests to obtain audit evidence to detect material misstatements in the financial statements. They are generally of two types: (a) analytical procedures; and (b) other substantive procedures, such as tests of details of transactions and balances, review of minutes of directors' meetings and enquiry'* (Glossary of Terms to the Statements of Auditing Standards: 1995). This is a less strong statement than that of AG307 (1989), since it is harder to gather evidence to prove something is correct to some close tolerance, than it is to obtain evidence of potential misstatement. The Auditing Standards also make clear that some form of substantive test must be done for all significant – that is material – amounts in the financial statements: '*Regardless of the assessed levels of inherent and control risks, auditors should perform some substantive procedures for financial statement assertions of material account balances and transaction classes'* (SAS 300.8). Thus substantive tests are an important part of the auditor's evidence.

In the context of this thesis, the standards make it clear that APs are one form of substantive test: '*Analytical procedures may also be performed as substantive procedures designed to obtain audit evidence directly*' (paragraph 4 of SAS410). That guidance is followed by auditors. In the survey described in Chapter 6, it was seen that if the systems of financial control are judged to be good, then the substantive test was likely to be some form of AP. This finding was supported by the work of Mulligan and Inkster (1999), who reported that APs are the sole source of substantive evidence in at least 29% of all audits. The Auditing Standards allow this: '*The assessed levels of inherent and control risks cannot be sufficiently low to eliminate the need for auditors to perform any substantive procedures for material account balances and transaction classes. However, these substantive procedures may comprise only APs where such procedures provide sufficient appropriate evidence*' (SAS 300). Thus the Standards recognise that APs may be the only substantive evidence obtained by the auditor.

9.3 How powerful should the substantive test be?

The legal case discussed in Chapter 4 indicated that the Law Courts expect the auditor to design the audit procedures to be such that, if material error exists, the procedures will have a reasonable expectation that it will reveal that error. In addition to the legal presumption, the SAS's give some indication that a substantive test should be powerful enough to indicate that material – significant – error might exist. Consider SAS 300 where the Standards recognise that there is a risk that the substantive procedure fails to detect a material error. The SASs call such a risk 'Detection Risk' and says that: *'Detection risk is the risk that auditors' substantive procedures (tests of details of transactions and balances or analytical procedures) do not detect a misstatement that exists in an account balance or class of transactions that could be material, either individually or when aggregated with misstatements in other balances or classes'* (SAS300). The SAS goes on to say that: *'Misstatements discovered in conducting substantive procedures may cause auditors to modify the previous assessment of control risk'* (SAS300, paragraph 52). Taken together these imply that a substantive test – i.e. an Analytical Procedure – should be powerful enough to detect material misstatement, or powerful enough to modify earlier assessments.

Now consider paragraph 10 of SAS 400, which says: *'In seeking to obtain audit evidence from substantive procedures, auditors should consider the extent to which that*

evidence together with any evidence from tests of controls supports the relevant financial statement assertions' (Paragraph 10 of SAS 400.3). As a consequence auditors might argue that it is the total of the evidence that is important, and that the purpose of a minimal AP is to provide extra comfort that all is well. Such an argument misses the point that at each stage the procedure should be capable of signalling the potential for a material misstatement. Each piece of evidence adds to the overall picture, and it is when all the stages are taken together that they provide sufficient evidence that there is no material misstatement (error). However, each should have the potential to identify, or at least signal, potential material misstatement or error.

For example, if in the auditor's judgment the management have a good grasp of their business, that there are adequate controls and the residual risk is low, then as a consequence they may decide to undertake some minimal level of substantive testing, as required in the Standards. In this example, that level of testing should be powerful enough to indicate if the prior assessment is incorrect, and that the management controls are not as good as assumed.

9.4 The evidence from the surveys

In order to find out more about the application of APs and to establish that APs are regularly used and are an important source of audit evidence, a number of surveys into their use were considered and a new survey was carried out. The surveys revealed several common themes. All the surveys confirm that APs are used a great deal. Also their use has increased since the 1980s and they provide a significant proportion of the total substantive audit evidence. Simple APs such as trend analysis, ratio analysis and scanning predominate: possibly using high-level data. So clearly APs are an important source of substantive audit evidence.

That APs are regarded as important is re-enforced by the auditor's perception, that APs first signal over 40% of all errors discovered during an audit. This was reported in Chapter 5. For example, in their survey, Biggs and Wild (1984) found evidence that APs initially signalled 42% of all errors, and subsequent surveys have confirmed that figure. It is important to note that the surveys, described in Chapter 5, were dealing with auditors' perceptions. In contrast, research by Bell and Knechel (1994) analysed documented audit files and found that APs produce about 16% of all the error signals. It matters not, in this context, whether the level of discovered error is the auditor's

perception of errors found or whether the small sample of actual audit files provides the levels of errors signalled by APs. Either figure implies that APs are important.

Another finding, reported in Chapter 5, was that APs are perceived by auditors to be effective, which Fraser, Hatherly and Lin (1997) surmised to mean cost-effective. Their interpretation is supported by the survey reported in Chapter 6, where auditors considered APs to be inexpensive to apply.

The survey carried out for this thesis, and reported in Chapter 6, confirmed the findings of the other surveys and extended those surveys by looking at quality issues. The new survey found that many of the audit firms did not appear too concerned about the quality of the data or of the rigour of the calculations. This statement is consistent with the view of many of the firms that APs are inexpensive. If more attention were paid to the data, APs might cease to be inexpensive procedures. For example, the use of data with a greater degree of disaggregation might produce more precise results, but would be more costly. Or, if the data were checked, then that would add to the costs.

In section 6.3.3 it was seen that many auditors fail to define an outcome-value (C) for the outcome-rule $| \hat{Y}_T - Y_B | \leq C$. They said that if the expectation value, $\hat{Y}_T$, was close to the book value in the financial statements being audited, then that would be '*the right answer*'; but they did not define close in advance of their audit test. In section 6.3.3, the finding was reported that there was little or no consideration of explanation effect phenomenon. This, together with the willingness observed in Chapter 4 (Boys (1997)), for auditors to accept management representations too easily, means that there is a potential for differences to be explained away without a sceptical appraisal of those representations.

A concern of two of the interviewed audit firms was that auditors did not sufficiently understand the business in order to carry out substantive APs. The need to appreciate the inter-relations between data from different parts of the entity, or similar entities, underlies APs. As the Auditing Standard on APs says, one of the factors in the application of APs is: '*the knowledge gained during previous audits, together with the auditors' understanding of the effectiveness of the accounting and internal control systems and the types of problems that in prior periods have given rise to accounting adjustments*' (SAS 410).

Also, in the survey carried out for this thesis, there was an example of auditors demonstrating that there is not always a link between their decisions and the consequences. Many of the respondents considered the nature of the substantive evidence they needed to collect, but they did not always appear to consider the extent of the substantive evidence that was required. One example given was that they might choose to carry out a sampling procedure, but the size of the sample was not linked to other criteria such as materiality, but was, instead fixed at an arbitrary upper size of 30. That indicates a lack of understanding of the procedures they apply.

So it is possible to conclude that the observations of Humphrey and Moizer (1990) and Garvey and Dietz (1997) that auditors did not understand the procedures they used, still has some relevance today. In particular, when the findings are taken together, they point to a lack of structure to the AP process, or at least to parts of it. If auditors understood the reasoning/theory that underpins the APs, one would not have expected these lapses. The statement by one of the Technical Partners interviewed in the survey about the desire to have flexibility to use their audit judgement says it all: '*The use of statistical type methods impaired and restricted my audit judgement*'. This view of statistics being a series of formulae and not a rationale for analysing data may be the reason for this statement. Certainly the auditor should be demonstrating a rationale to the structure of their AP work. Currently, based on the evidence gathered in the new survey, there appears to be little structure to the use of APs, certainly none that is articulated, or documented, in any way.

9.5 APs and the Business Risk Audit Methodology

Most of the surveys dealt with audits conducted under a methodology where the Audit Risk Model philosophy underpins the logic for combining information. To paraphrase Lemon, Tatum and Turley (2000), the Audit Risk Model has been a basic fundamental of audit methodology for the last 15 to 20 years. That approach is still applied by most of the firms outside of the big five (or big four as they now are). But for firms who deal with the audit of global entities a new approach has been developed and implemented. It is the Business Risk Audit Methodology (BRAM).

9.5.1 The Essential parts of Business Risk Audit Methodology

Before considering the impact of the BRAM methodology on the use of APs, it is pertinent to consider a motivation for that methodology. Bell, Marrs, Solomon and Thomas (1999) claim that important questions are left unanswered by the old audit approach. For example, '*what level of business knowledge is needed? How much knowledge is needed to obtain reasonable assurance that the assertions in the financial statements are not materially misstated? And what is the relative importance of this business knowledge to the auditor's opinion*'? Such arguments apply to many parts of the SASs, which are deliberately at a high level, allowing the auditor to interpret them into the audit strategy of each firm. This 'understanding of the business' is given as one reason for the proposed new approach. However the concerns about business knowledge are echoed by the findings of the survey described in Chapter 6: there some firms were worried about a lack of understanding of the business that is fundamental to APs and, one would have thought, to any financial audit. So it may be that more specific audit guidance is needed, rather than, or as well as, the new/revised approaches.

Another motivation may be found in a perceived threat to the audit profession from analysts who have gained access to the databases of all the information of an entity, and who no longer have to rely on the summary of the information that makes up the financial statements. With powerful computers the analyst can repeat the analysis many times in the year, and therefore does not have to rely on the annual snapshot: the financial statements. Elliott (1994) put it this way: '*Early in the century* (the 20th Century), *financial statements represented a large part of the information available to an enterprise's debt and equity investors. As accounting principles improved, the value of the financial statements also improved. But, facilitated by information technology, other sources of relevant information are increasing available; for example, investors can get up-to-the-minute data about companies through public and proprietary databases without waiting for quarterly or annual reports*'. However, this misses one important point. The annual statements have an added value since they are supposed, after independent audit, to be attested to their 'correctness' – the truth and fairness of those statements. That said, auditors were probably worried about the loss of providing

additional services over and above the financial audit work to certify the financial statements.

Evidence that methodologies such as BRAM are in use by one of the major audit firms, reflecting the move to a more business orientated audit approach, was noted during an interview for the survey reported in Chapter 6. It was with one of the major audit firms in the USA. The underlying business audit model was discussed and copies of the 'partner cards' for the approach were provided to illustrate its application and the use of APs as a part of the audit methodology. That was the second interview with that particular firm. An earlier interview took place in 1996 in the UK, and it was subsequently clear that the audit philosophy discussed in the USA was not an entirely new one, but a natural evolution of their thinking. Fundamental to both approaches were minimal substantive tests if the business systems were judged to be capable of preventing material error. Those tests are likely to be APs.

For the purpose of conducting a financial-statement audit using the BRAM approach, the key feature is for the auditor to gain an understanding of a client's business. In particular the auditor directs their attention to the client's systems dynamics – its strategic positioning within its environment, its emergent behaviours that impact on the attained level of performance, the strengths of its connections to outside economic influences, the specific interrelationships and internal processes interactions that dominate its performance, and potential changes from outside that might threaten the viability of the client's strategies and niches. Bell, Marrs, Solomon and Thomas (1999) concluded that: *'Through the application of strategic-systems-oriented analytical procedures and the related knowledge-assimilation process, the auditor gathers, analyses, and integrates reliable and independent information which becomes expert knowledge. Such expert knowledge, in turn, enables the auditor to reach valid and defensible conclusions about the appropriateness of financial statement assertions'*. At the conclusion of the BRAM approach the auditor uses a reduced level of substantive test (reduced from previous methodologies) to direct the final audit and to obtain evidence to meet the residual risks. Eilifsen, Knechel and Wallage (2001) put the philosophy this way: *'Because of the extensive knowledge acquisition (evidence) used to assess business risks and their links to audit risks, only limited substantive evidence*

may be necessary for many assertions/objectives'. Such substantive tests are likely to be APs.

Effective APs require a good understanding of the business as the Auditing Standards acknowledge. Bell, Marrs, Solomon and Thomas (1999) put it this way: '*APs, by their very nature, are complex and require an adequate understanding of business activities and other underlying economic phenomena*'. Such knowledge should produce better investigation-rules: by better it is meant that the rules more closely model or replicate the relationship being audited to produce an audit expectation. This then brings us back to deciding what type of AP to use. If the auditors are still using simple APs, such as ratios, ratio trend analysis and common-size financial statements – see page 61 of Bell, Marrs, Solomon and Thomas (1999) – then the issue is whether these APs are powerful enough to overturn, or at least indicate a problem with earlier audit judgements, based on assessments of management and their systems.

9.5.2 Business Risk Audit Methodology: Additional services and Audit Independence

Despite bringing benefits to the audit through a more comprehensive understanding of the audit client, the BRAM approach raises major worries in relation to the motivation for this audit process and the potential independence of the auditor. These impact directly on the AP process. Bell, Marrs, Solomon and Thomas (1999) assert that the BRAM methodology will facilitate the provision of services over and above the basic audit. They claimed that: '*Also the client business model will provide a basis for the auditor's provision of additional assurance to management about the entity's achieved level of performance in terms of its operational, financial reporting and compliance objectives and goals*' (author's underlining).

That part of the purpose of the BRAM process was to get on side with management was clearly described in the ICAEW report by Lemon, Tatum and Turley (2000). They found that: '*In some cases the firms are explicit in suggesting that one purpose of the business risk approach is that it assists the auditor to form a view of the entity that is consistent with that of management. As stated by one firm, the new approach is useful for 'aligning our audit process more closely with managements needs to improve business performance and manage business risks'. In the case of one firm, this emphasis was reflected in the fact that part of audit approach allowed for management*

to influence the specific engagement objectives for the annual audit, over and above the auditor's professional responsibilities, and to set key 'deliverables' that the audit firm would seek to fulfil and would explicitly be held accountable for at the end of the audit cycle'.

The emphasis on the opportunity for providing an additional service to the audited client, as shown by the underlining in the above statement by Bell, Marrs, Solomon and Thomas (1999), raises a number of questions that are relevant to the use of APs within the BRAM methodology. Such as, who is the real client: is it the senior management, or the shareholders or other stakeholders such as potential shareholders? And, is there a danger of the financial audit and the client services getting confused? Another question, post Enron and Worldcom, is the issue of auditor independence, and do such methodologies as BRAM threaten to compromise the independence of the external auditor? For example, in subsequent audit periods, will the auditors be auditing their own work and/or guidance? This is relevant since APs rely on the auditor's integrity. Their APs may use the data from the client. They have to seek explanations from the client about differences between the audit-estimates and the figures in the financial statements, and they have to be aware that the AP may be an important source of evidence to refute their assumptions about the audited client.

The new methodologies, which appeal to the rationale that it must be a good thing if auditors better understood their clients, are unproven, whereas the older methodologies have identified weaknesses. Eilifsen, Knechel and Wallage (2001) counselled some caution about the new approach: '*There is some understanding of the limitations of traditional methods due to previous research; a similar level of experience and insight must be attained before we can similarly assess the new audit methods*'. Thus the new approach needs to be monitored and researched to assess whether its intrinsic appeal is translated into a methodology that produces evidence to support the auditor's opinion on the truth and fairness of the financial statements.

9.5.3 Business Risk Audit Methodology and Substantive Tests

Where this consideration of BRAM is particularly relevant to the effectiveness of APs is when substantive testing is reduced to a minimum. The issue is then whether such procedures will detect material error, should it be present, rather than the presumption made by auditors that such error does not exist. Eilifsen, Knechel and Wallage (2001)

pointed out that '*to the extent that traditional substantive testing is decreased or eliminated, an auditor runs the risk of overlooking errors that might have been detected under those approaches*'.

There is not necessarily universal agreement within the profession that such new methods are relevant to all audits. A comment by a partner in a major audit firm after the introduction of the BRAM approach was: '*For companies that do not need an audit, this approach is as good a way of not doing an audit as any other*'. While such scepticism may be motivated by the distrust of change, it can represent a healthy caution that the new methodologies may not be any better than those they replace.

The introduction of the BRAM approach, which is only likely to be widely used by the auditor of larger complex organisations should, in theory, produce better APs. The structure being advocated earlier in the thesis – the investigation-rules, outcome-rules and explanation-rules – ought to be more likely because of a better understanding of the business. This should ensure that the auditor considers the potential shortcomings of the proposed APs. For example if Bell, Marrs, Solomon and Thomas (1999) are correct, the BRAM will '*serve to heighten the auditor's level of professional scepticism by enabling the auditor to develop his own hypothesis about unusual financial statement trends and fluctuations without relying exclusively on managements' explanations*'. Such a conclusion is entirely consistent with the earlier arguments about the explanation effect, and would help to address the concerns raised in Chapter 6 about that effect. A better understanding will enable the auditor to form the alternative explanations in the way discussed in section 4.10. However, only further research will determine whether Bell, Marrs, Solomon and Thomas's (1999) hypothesis is achieved in practice.

The use of APs is likely to continue to be significant, a view that is consistent with the results of past and recent surveys, such as Fraser et al (1997), Mulligan and Inkster (1999) and the survey for this thesis. Indeed Lemon, Tatum and Turley (2000) found in their interviews with the large firms that: '*the movement to a greater emphasis on APs as a source of evidence for the financial statements continues under business risk auditing*'. They also found that: '*While all of the firms generally require performance of some substantive procedures and where risks are high, a commensurate level of*

evidence will have to be collected, in low risk situations detection risk can be addressed primarily by APs performed in connection with the investigation of business risk'.

Thus it is apparent that under either audit approach the level of substantive procedures is diminishing and that such procedures are likely to be APs. So it is vital that such APs should produce sufficient evidence to confirm or deny the auditor's judgement about the organisation.

9.6 How do APs perform: the experiments described in Chapters 7 and 8

Having established that APs are an important substantive procedure, and that simple APs dominate, Chapter 7 (the experiments of others) and Chapter 8 (the experiments conducted for this thesis) then addressed the issue of whether APs, and simple APs in particular, were effective. That is to demonstrate if APs, as the only substantive test, really fulfil their purpose at the substantive stage. In these experiments effectiveness was measured by the levels of false-negative and false-positive signals for different levels of outcome-value and seeded error. False-negative signals imply that the investigation-rule is detecting error that is not there, and false-positive signals are missing material error that is present.

The types of investigation-rules used in Chapter 8 were grouped into two classes. The first was of a regression type, similar to those used by Knechel (1988a) and Kaplan (1978), while the other set were simple rules similar to those used by Loebbecke and Steinbart (1987) and Knechel (1988b). The latter type of simple investigation-rules are typical of those used extensively by auditors. The investigation-rules were summarised in Table 8.2.

9.6.1 The data used for the experiments for this thesis and why it was chosen

To investigate whether or not the chosen investigation-rules would have a low propensity for false signals, they were applied to three different populations of monthly data. There were 2,500 sets of data simulated for each of the three populations.

One set of data was simulated to be similar to the data created by Knechel (1988a): the data was increasing in a structured way and the random error was distributed as if it came from a normal distribution. The reason for using this data was to assess how the regression based investigation-rules used by Knechel would perform, when applied to data that was similar to his and when the criteria for effectiveness was to measure

directly the false signals rather than the reduction of other tests measured through the Audit Risk Model.

The other two sets of data were simulated to have the characteristics of monthly pay data from two Government Agencies. These data were chosen to see how different investigation-rules would perform on data that was from what is, for many organisations, a well controlled, stable area: that is the pay of the staff. It should be noted that the pay data was from the period 1990 to 1994/5, before the change, on a large scale, within the public sector to performance-related pay. It is likely that there will be more variations in pay data from the latter part of the 1990s. Thus the earlier period represents an ideal data set to test APs, in the sense that APs would be expected to work well here, where there is a strong relationship. If they do not, that will be valuable information because it will demonstrate the potential failings of APs when they could be expected to do well.

Thus the experiments described in Chapter 8 used real data, which many of the earlier experiments did not, and data that was similar to Knechel's so that the experiments could directly test his results by trying to replicate his work but without the, flawed, Audit Risk Model to measure the effectiveness. The use of the two sources of data, similar to the Agencies and to Knechel's, provided the opportunity to compare the performance of investigation-rules with different types of data.

9.6.2 Regression type APs used by Knechel and the experiments carried out for this thesis

The first group of investigation-rules were the regression type, similar to those used by Knechel (1988a). They were applied to monthly data and monthly outcome-rules were used to measure the results of the experiments. It might be thought that regression type APs would perform very well, however Kaplan (1978) found that this was not uniformly true. He reported on an experiment using APs based on regression in which the expectation value never got closer than within 10% of the book value. However, other analysts using regression or time series based APs did not report such problems about the effectiveness of APs. Their experiments were reported in section 7.7 and they included the work of Knechel (1988a) and Ameen (1989). They evaluated the effectiveness of the APs by reference to the extent of the reduction of another substantive procedure: sampling. The Audit Risk Model was used to calculate the

revised substantive sample sizes. This reduction does not necessarily provide support for the effectiveness of such APs because of the defects of the Audit Risk Model. The extent of the reduction may be exaggerated or underestimated. Thus the evidence from these studies was not conclusive in demonstrating whether APs are effective.

In order to assess the worth of the AP, its performance must not be confused with, or disguised by, its combination with some other substantive audit procedures. Therefore, the experiments described in Chapter 8, used the criteria of whether there were low frequencies of false signals.

The results of the experiments described in Chapter 8, using the regression based investigation-rules and monthly data, showed that the regression-rules produced wide variations in performance. There were false signals for many combinations of outcome-value and seeded error. Also the levels of false signals were not consistent between investigation-rules (with the same data) or between data sets for the same rule.

Specifically for the data simulated to be similar to the two real populations, there were false-negative signals for many combinations of error and outcome-value when the error was spread over 1, 3 or 6 months. However, when the error was spread over all twelve months, then the signals were false-negative when the outcome-value (C) > the seeded error (E), and false-positive when C < E.

When the same simple regression investigation-rules were applied to data that was similar to Knechel's (1988a) they performed better than for the 'real' data when the error was spread over 1, 3 or 6 months: they produced lower levels of false-negative signals. However when the error was spread over 12 months, there were false-positive signals. This is significantly different to the results of Knechel. It is likely that his use of the Audit Risk Model disguised the occurrence of false signals, and thus his finding that APs were effective may well be flawed.

This finding, together with the inferior performance when the investigation-rule is used in conjunction with the Agency data, points to a conclusion that regression based investigation-rules do not perform well with data of good quality. Such data may even be better than the data that auditors typically use, and therefore APs used in conjunction with poor quality data may not perform well.

To sum up, there were levels of false signals that indicate such APs may lead to additional work, or more seriously, in certain circumstances, provide false-positive signals. The implications of this are taken up in section 9.7.

9.6.3 Simple APs used by Loebbecke & Steinbart and in the experiments carried for this thesis

In addition to the experiments described in Chapters 7 and 8 that investigated the effectiveness of some APs that used regression type investigation-rules, there were also experiments using 'simple' investigation rules: comparison, trend and ratio analysis.

In Chapter 7, the work was reported of Loebbecke and Steinbart (1987). They found that many of the simple APs are not effective, because such investigation-rules had high frequencies of false-negative and false-positive signals. This was true for outcome-values of 5%, 10%, and especially 15%. As well as performing badly, the APs were not consistent in their performance: that is a particular investigation-rule might perform well in one account area, and badly elsewhere. Another piece of research by Knechel (1988b) looked at the application on non-statistical APs. His work, discussed in section 7.8, used twelve investigation-rules based on ratio and trend models and were applied to the same data as Knechel (1988a). The effectiveness of the Knechel (1988b) investigation-rules was measured by how much the rules reduced the sample sizes, as with the Knechel (1988a) research.

To build on that work and to investigate whether the findings of Loebbecke and Steinbart (1987) or Knechel (1988b) were correct, in Chapter 8 a series of experiments were carried out. The seven experiments applied investigation-rules 1.3 to 1.9 of table 8.2 – experiments five to eleven – to the agency data, described in section 9.6.1, after it had been summed into annual totals. That is, four annual amounts rather than the forty-eight monthly totals. Here the rules were applied to good quality data that was, probably, more consistent than that used by Loebbecke and Steinbart (1987). One reason for this was that their data came from a period of high inflation.

Sections 8.9.1 to 8.9.4 set out the detailed results for the simple investigations-rules. In summary, while the outcome-value (C) and error (E) were close, there were false signals: false-negative while $C > E$, and false-positive while $C < E$. Close can mean an absolute difference of up to 3% of the value being tested. These results are entirely

consistent with the work of Loebbecke and Steinbart (1987) who found high frequencies of Type I and II error, the equivalent of our false-negative and false-positive signals. This is, at first sight, surprising since the new data was well controlled, whereas the Loebbecke and Steinbart (1987) data was from a period of high inflation. One conclusion is that such investigation-rules perform badly whether the data is of a good quality or not.

The combined results of the simple investigation-rules imply that the conclusions of Knechel (1988b) are also likely to be incorrect. His conclusion that the sample sizes were reduced is not always supported by the experiments in Chapter 8, where the results clarify the shortcomings of the APs for some combinations of outcome-rule and error.

9.7 Summary of the conclusions and their implications for the auditor

It has been demonstrated that APs are regarded to be an important procedure whether the philosophy of the audit methodology is the Audit Risk Model or the Business Risk Audit Methodology. The APs are a significant proportion of the substantive evidence. In addition it has been shown that the APs analysed in this thesis, used with stable data from well controlled sources, gave high levels of false-negative signals while the error in the audited value was low and, if the outcome-value was typically 4% or less of the value being tested, this varied with the investigation-rule and the data source. The implication for the auditor is that, in such circumstances, the signal they receive is that additional work is needed. If that signal were wrong that would mean that more (nugatory) work is required, as, although it would increase their knowledge of the audited organisation, it would be unnecessary.

For the annual outcome-rules, particularly when used with simple investigation rules there were false signals, while the outcome-value was close to the seeded error. Those were false-negative signals when the outcome-value was greater than the error, with the implication explained above. When the outcome-value was less than the error, and the outcome-value and the error were close, there were false-positive signals.

Such false-positive signals were also observed with monthly outcome-rules where the seeded error was evenly spread over the audited period. If in these circumstances (the false-positive signals) the AP is being used because the auditor has decided, based on their assessment of management systems and other audit work, that a minimal

substantive test is all that is required. Then in these circumstances, unless the error is well in excess of the outcome-value, the auditor would not be aware that they should be querying their earlier decisions. The auditor's assumptions about the acceptability of the assertions in the financial statements, based on management controls and their understanding of the business, will not be challenged by the minimal substantive tests – APs.

The auditors interviewed in the survey, reported in Chapter 6, did not report high frequencies of false-negative signals. One possible conclusion from that, together with the results of the experiments in Chapter 8, is that auditors may be using large outcome-values for each of the different account areas within a set of financial statements being tested. In this instance 'large' is relative to the value being audited by each particular substantive test. If large outcome-values are used then there is a question about the combined effect of these and whether it exceeds the material amount set for the financial statements as a whole. Over the last few years, there have been discussions about the 'Expectation Gap', which has been described as: '*a representation of the feeling that auditors are performing in a manner at variance with the beliefs and desires of those for whose benefit the audit is being carried out*' (Humphrey (1997)). The materiality to which auditors work, and any links between that and the outcome-values, is a part of this expectation gap. If the users of the financial statements suspected that the balances within the financial statements were checked to within a value in excess of 10% or even 5% of those balances, would they be happy? The relationships between the setting of audit materiality and of local outcome-values are an important subset of the combination of audit evidence. This is an area that was not explored in this thesis.

The move to new methodologies may be useful in the sense that auditors will develop better investigation-rules and explanation-rules. But an implication from the work of this thesis is that auditors may be using relatively high outcome-values, as that is one way of avoiding high levels of false signals. A further piece of research could be to consider the link between those outcome-values and the overall materiality.

Another issue identified in earlier work and the survey carried out for this thesis is an understanding of APs, or at least an articulation of that through the actions of the auditor. Although Bell, Marrs, Solomon and Thomas (1999) put forward the

proposition that the BRAM philosophy should improve audit work through a better understanding of the organisation, another step is also required as part of the strategy. It would apply to any audit requiring substantive evidence: be it under the Audit Risk Model or Business Risk Audit Methodologies. It consists of providing a simple structure for APs as discussed in section 3.8, and added to in section 4.10. That structure can be split into three rules: namely an investigation-rule, an outcome-rule and an explanation-rule. This is necessary, as can be seen from Chapters 5 and 6 where it is shown that such rules are not generally used.

The purpose of the investigation-rule is to explain, and make transparent, the causal links between different sets of data. It utilizes the auditor's understanding of the plausible relationships between items in the financial records of the audited entity. The output from this is some form of expectation value: $\hat{Y}_T$. The next stage is to assess the result of applying the investigation-rule and to measure whether the difference between the expectation-value ($\hat{Y}_T$), and the value being tested (Y_B) is acceptable. This is the outcome-rule and it should be less than some critical value called the outcome-value (C): $|\hat{Y}_T - Y_B| < C$. The final rule is the explanation-rule. Its purpose is to ensure that the auditor obtains a valid reason for, or otherwise substantiates, the discrepancy between $\hat{Y}_T$ & Y_B. A suggested structure for explanation-rules was set out in section 4:10, this was:

1. Identify potential alternative explanations before doing the AP;
2. Carry out the AP;
3. Review the alternative explanations before talking to the audited organisation;
4. Seek explanations from the audited organisation; and
5. Substantiate those explanations.

These three rules do not overcome the potential for high levels of false signals, but they make the auditor aware of potential issues, and can demonstrate that awareness to others.

This set of rules is entirely consistent with the strategy of BRAM as described by Bell, Marrs, Solomon and Thomas (1999). None of the proposed rules need to be bureaucratic since, for simple investigation-rules, a short description of the process

would be sufficient. However, it would make it clear what has been done to the person reviewing the work, as well as for the auditor performing the work.

To sum up, APs are used extensively, simple techniques are the most common and auditors perceive that APs are effective and have categorised them as cost effective. The surveys described in Chapters 5 and 6 imply that APs may be low cost, but the evidence from this thesis is that in certain circumstances they may be ineffective rather than cost effective.

9.8 Areas for further research or further action

Following on from the findings of this thesis there are some possible areas for further research. These are set out below:

(1) Auditors appear to be using large outcome-values for each of the different account areas within a set of financial statements being tested. There is a question about the combined effect of these outcome-values and whether it exceeds the material amount set for the financial statements as a whole.

There are two pieces of research that could flow from this: [A] a survey and [B] the investigation of a rationale for combining the outcome-values.

> [A] Auditors, if they do not use large outcome-values, are likely to be getting high frequencies of false-negative signals, thus causing nugatory work. A new survey would be useful to determine if auditors do get such nugatory signals, and to ascertain how those signals are addressed.
>
> [B] Associated with [A] is a rationale for combining the results of different audit procedures. It would assist the auditor in their judgements about the financial statements.

(2) It has been shown that many investigation-rules have a potential for false-negative and false-positive signals. The problems might be overcome if the investigation-rule was based on non-parametric regression. Although some work on this subject was done in the development of this thesis, because of time constraints, it was not pursued. A further exploration of the usefulness of such APs could identify whether they could efficiently and effectively assist the auditor. The work should be associated with a programme involving auditors to see if such methods could be, and would be, acceptable to them.

(3) It would also be useful to repeat the experiments of this thesis with data from other areas within the financial statements and / or different outcome-rules, possibly an annual rule in place of the monthly rules.

(4) It is now some time since the survey for this thesis was completed. Audit is a rapidly developing process. Further research into the impact of the SASs since they were issued in 1995 might find that auditors have addressed some of the issues raised in this thesis.

(5) An AP is only one of many signals that the auditor uses in the audit process. When it is the only substantive evidence, false-negative signals may cause extra work, but false positive signals pose a particular risk. Properly done APs are a powerful tool. Indeed at the planning and the review stages they are mandatory. In the spirit of the IDA philosophy of Chatfield (1985), they provide a very useful analytic tool and assist in an understanding of the business being audited. However, auditors do not appear to be completely aware of the potential hazards of any predictive procedure, such as the need to verify data, or the imprecision associated with prediction. One way of overcoming this problem may therefore be that the APB should issue an Audit Bulletin drawing attention to the strengths and weaknesses of this potentially important audit tool: APs.

Appendix A

The skeleton questionnaire used the survey described in Chapter 6.

Name of the firm:
Partner/ Manager:
Date:
Time:
Contact Points:

AREAS TO BE COVERED:

A DEFINITION OF APS

- What do you define APs to be?

B USE OF APS BY FIRM

What APs do you use?

- Graphical methods
- Plotting
- IDA
- Trend Analysis
- Ratio Analysis
- Time Series
- Regression
- Other techniques

At what stages of audit do you apply APs

- Planning / substantive / review

C WHAT SOURCES OF DATA ARE USED IN APS

- Solely comparing internal data
- Is there a requirement to use independent data

D WHAT QUALITY CHECKS ARE DONE ON THE DATA USED IN THE APS

- What quality control checks do you do on the data?
- Do you review/check the systems producing the data used for the APs?
- Is there a requirement to do the above?

E HOW DO YOU COMBINE WITH OTHER EVIDENCE?

Including how the evidence combined with other substantive evidence.

F EXPLANATION OF VARIANCES

- Requirement to explain variances?
- What guidance is given on obtaining evidence to explain variances?
- How do you cope with the explanation effect?
- Review procedures?

G DOCUMENTATION

- How do you document?
- What are the standards?

H COMPUTER TOOLS

- Do you use computer packages?
- Do you use statistical packages?
- Do you use computer tools?
    - Spreadsheets
    - Database packages
    - Statistical packages
    - Time-series packages
    - Bespoke packages
    - Other?

I AUDIT PRACTICES COMMITTEE

- What developments are there likely to be?
- Did new SAS change your application of APs?

J GENERAL

- Why firms use APs?
- Why does your firm use APs?
- Any doubts/worries/concerns about the profession's use of APs?
- Any doubts/worries/concerns about your firm's use of APs?
- Failures?

BIBLIOGRAPHY

Albrecht, W. S. and McKeown, J. C. 1976. *Towards an Extended Use of Statistical Analytical Reviews in Audit.* Symposium II in Illinois, p. 70-79.

Aldersley, S. J. 1989. Discussion of Achieved Audit Risk and the Audit Outcome Space, *Auditing: A Journal of Practice and Theory,* 1989, vol. 8, Supplement, p. 85-97

Altman, E. I. 1968. Financial Ratios, Discriminant Analysis and the prediction of Corporate Bankruptcy, *The Journal of Finance*, 1968, vol. XXIII, no. 4, p. 589 - 609.

Ameen, E. C. 1989. *The Effectiveness and Efficiency of Statistical Analytical Review in an Auditing Strategy: A Comparison of Regression and the X11 Model.* PhD Thesis at University of South Carolina.

Ameen, E. C. and Strawser, J. R. 1994. Investigating the use of Analytical Procedures: An Update and Extension, *Auditing: A Journal of Theory and Practice,* Fall 1994, vol.13, no. 2, p. 69-77.

American Accounting Association, 1973. *Committee on Basic Auditing Concepts.* New York.

American Institute of Certified Public Accountants. 1988. *Statement on Auditing Standards Number 56, Analytical Review Procedures.* New York: ACIPA.

American Institute of Certified Public Accountants. 1998. *Auditing Practice Release on Analytical Procedures in 1998,* New York: ACIPA.

Anderson, C. A. and Sechler, E. S. 1986. Effects of Explanations and Counter explanations on the Development and Use of Social Theories. *Journal of Personality and Social Psychology,* 1986, vol. 50, No. 1, p.23-34.

Anderson, R. 1977. The Interrelationship of Compliance and Substantive verification in Auditing, In *Cushing and Krogstad (Eds.), Frontiers of Auditing Research,* Bureau of Business research, University of Texas at Austin, 1977, p. 63 – 137.

Andrews, R. and Smith, T. M. F. 1983. Pseudo-Bayesian and Bayesian Approach to Auditing. *The Statistician,* 1983, vol. 32, p. 124- 126.

Andrews, R. and Smith, T. M. F. 1987. *A Bayesian Model of the Audit Process,* Unpublished paper.

APC – AG 307, 1989. *Banks in the United Kingdom,* Audit Practices Committee, March 1989.

Barnes, P. 1982. Methodological Implications of Non-normality Distributed Financial Ratios. *Journal of Business Finance and Accounting,* 1982, vol. 9, no.1, p. 51 - 62

Barnes, P. 1987. The Analysis and Use of Financial Ratios: A Review Article. *Journal of Business Finance and Accounting,* Winter 1987, vol. 14, no. 4, p. 449 - 469.

Beattie, V., Brandt, R and Fearnley, S. 1997. Look To Your Laurels. *Accountancy,* July 1997, p. 140 - 141.

Bell, T. and Knechel, W. R. 1994. Empirical Analyses of Errors Discovered in Audits of Property and Casualty Insurers. *Auditing A Journal of Practice and Theory,* Spring 1994, vol. 13, no 1, p. 84 - 100.

Bell, T. B., Marrs, F.O., Solomon, I. and Thomas, H. 1999. Auditing Organisations through the Strategic-Systems Lens. The KPMG Business Management Process. *KPMG – University of Illinois Monograph*

Berry, R. H. and Nix, S. 1991. Regression Analysis versus Ratios in the cross-section analysis of Financial Statements. *Journal of Accounting and Business Research,* 1991, vol. 21 no. 82, p. 107 - 118

Biggs, S F and Wild, J J. 1984. A note on the practice of Analytical Review. *Auditing: A Journal of Practice and Theory,* Spring 1984, vol. 3, no. 2, p. 68-79.

Blocher, E and Patterson Jr, G F 1996. The Use of Analytical Procedures. *Journal of Accountancy.* February 1996, p. 53 – 55.

Boys, P. (1997). *Raising Auditing Standards: The Impact of DTI Inspectors' Reports 1971-1995*. The Canterbury Business School.

Burton, J. C. and Fairford, P. 1982. Auditing evolution in a changing environment, *Auditing: A Journal of Practice and Theory*, Winter 1982, vol. 1, no. 2, p. 1- 22.

Carmichael, D. 1968. Tests of Transactions - Statistical and Otherwise. *Journal of Accountancy*, Feb 1968, p. 36 – 40.

Chen, Y and Leitch, R A. 1998. The Error Detection of Structural Procedures: A Simulation Study. *Auditing: A Journal of Practice and Theory*, Fall 1998, vol.17, no. 2, p. 26 - 78.

Chatfield, C. 1975. *The Analysis Of Time Series: Theory And Practice.* Chapman and Hall, 1975.

Chatfield, C. 1985. The initial analysis of data (with discussion). *J.R.Stat. Soc. Series A*, 1985, vol. 148, Part 3, p. 214-253.

Chatfield, C. 1995. *Problem Solving: a statistician's Guide.* Chapman and Hall Second edition 1995.

Coakley, J. 1982 *Analytical Review: A comparison of Procedures and Techniques Used in Auditing.* PhD thesis at University of Utah 1982.

Daniel, S. J. 1988. Some Empirical Evidence about the Assessment of Audit Risk in Practice. *Auditing: A Journal of Practice and Theory.* Spring 1988, vol. 7, no. 2, p. 174 - 181.

Daroca, F. P, and Holder, W. W. 1985. The use of Analytical Procedures in Review and Audit Engagements. *Auditing: A Journal of Practice and Theory.* Spring 1985, vol. 4, no. 2, p. 80-92.

Deakin, E. B. 1976. Distributions of Financial Accounting Ratios: Some Empirical Evidence. *The Accounting Review*. January 1976, p. 90-96.

Dzeng, S. C. 1991. *A Comparison of Autoregression Models and Ordinary Least Squares Multiple Regression models for Analytical Procedures.* PhD Thesis at Drexel University, 1991.

Eilifsen, A., Knechel, W.R. and Wallage, P. 2001. Application of the Business Risk Model: a Field Study. *Accounting Horizons,* Sept. 2001, vol. 15, no. 3, p. 193 – 207.

Elliott R K. and Rogers J. R. 1972. Relating Statistical Sampling to Audit Objectives. *The Journal of Accountancy,* 1972, p. 46 – 55.

Elliott R. 1994. The Future of Audits. *Journal of Accountancy,* Sept. 1994, no. 178, p. 74 - 82

Ezzamel, M. and Mar-Molinero C. 1991. *Initial Data Analysis of Financial Ratios: A fresh look at the Figures.* Discussion Paper 91 –13: University of Southampton Working Paper.

Flint, D. 1988. *Philosophy and Principles of Auditing: an Introduction.* Macmillan 1988.

Fraser, I. M., Hatherly, D. J. and Lin, K. Z. 1997. An empirical investigation of the use of analytical review by external auditors. *British Accounting Review,* 1997 vol. **29**, p. 35-47.

Garvey, H. and Dietz, A. 1997. What the JMU Finds. How are Firms coping with the new SASs? The JMU has ways of finding out. *Accountancy,* Sept.1997, p. 82 - 83.

Godsell, D 1991. *Auditors' Legal Duties And Liabilities.* Longman.

Hatherly, D. and Skuse, P. 1991. *Audit Reports, in M Sherer and S Turley, Current Issues in Auditing* (Second Edition: 1991), Paul Chapman Publishing, p. 115 – 130.

Humphrey, C. and Moizer, P. 1990. From Techniques to Ideologies: An alternative Perspective on the Audit Function. *Critical Perspectives on Accounting.* 1990 vol.1, p. 217-238.

Humphrey, C. 1997. *Debating Audit Expectations, in M Sherer and S Turley, Current Issues in Auditing* (Third Edition: 1997), Paul Chapman Publishing, p. 3 – 29.

Hylas, R. E. and Ashton, R. H. 1982. Audit Detection of Financial Statement Errors. *The Accounting Review* October 1982, vol. LVII, no. 4, p. 751-765.

ICAEW. 1990. Auditing And The Future: *Proceedings of an Audit Research Conference;* Edinburgh 11/12 April 1988.

Jiambalvo, J. and Waller, W. 1984. Decomposition and Assessments of Audit Risk. *Auditing: A Journal of Practice and Theory,* 1984, vol.3. no. 2, p. 80 – 88.

Kaplan, R. 1978. *Developing a Financial Planning model for an Analytical Review: A Feasibility Study.* Symposium III in Illinois: 1978, p. 3 – 34.

Kaplan, R. 1983. A Financial Planning model for an Analytical Review: The case of a Saving and Loan Association. *Auditing: A Journal of Practice and Theory,* Spring 1983, vol. 2, no. 2, p. 52 – 65.

Keynes, J. M. 1936. *The General Theory Of Employment Interest And Money,* Chap.24.

Kinney Jr,. W. R. 1978. ARIMA and Regression in Analytical Review: an Empirical Test. *The Accounting Review,* January 1978 vol. LIII, no 1, p. 48 - 60.

Kinney Jr., W.R. 1979. Integrating Audit Tests: Regression Analysis and Partitioned Dollar-Unit Sampling. *Journal of Accounting Research.* Autumn1979, vol. 17, no. 2, p. 456 - 475.

Kinney, W. R. and Salamon, G. 1982. Regression Analysis in Auditing: A Comparison of Alternative Investigation Rules. *Journal of Accounting Research.* Autumn1982, vol. 20, part 1, p. 350–366.

Kinney, W. R. Salamon, G. and Uecker W. 1986. *Computer Assisted Analytical Review, Accounting* Educational Monograph Number 7. American Accounting Association.

Knechel, W. R. 1988a. The Effectiveness of Statistical Analytical Review as a Substantive Auditing Procedure: A Simulation Analysis. *The Accounting Review.* January 1988, vol. LXIII, no. 1, p. 74 – 95.

Knechel, W. R. 1988b. The Effectiveness of Non-Statistical Analytical Review procedures used as Substantive Audit Tests. *Auditing: A Journal of Practice and Theory,* Fall 1988, vol. 8, no. 1, p. 87 - 107.

Koonce, L. 1992. Explanation and Counter-explanation in Audit Analytical Review. *The Accountancy Review.* January 1992, vol. 67, no. 1, p. 59 – 76.

Kreutzfeldt, R.W. and Wallace, W.A. 1986. Error Characteristics in Audit Populations: their Profile and Relationship to Environmental Factors. *Auditing: A Journal of Practice and Theory.* Fall 1986, vol. 6. no. 1, p. 21-43.

Law, A. M. and Kelton, W. D. 1991. *Simulation Modelling and Analysis.* McGraw-Hill, Inc. (Second edition).

Lemon, W.M., Tatum, K.W., Turley,W.S. 2000. *Developments in the Audit Methodolgy of Large accounting Firms*, ICAEW.

Leslie, D. A. Aldersley, S. J. Cockburn, and C. J. Reiter, 1986. *An Assertion-Based Approach to Auditing.* Auditing Symposium VIII. Lawrence, K A. ed. The University of Kansas.

Leslie, D. A. Teitlebaum, A.D. and Anderson, R.J. 1979. *Dollar Unit Sample: A practical Guide for Auditors.* Pitman Press.

Lev, B. (1980). On the use of Index Models in Analytical Reviews by Auditors. *Journal of Accounting Research.* Autumn1980, vol. 18, no.2, p. 524-550.

Lev, B. and Sunder, S. 1979. Methodological Issues in the Use of Financial Ratios. *Journal of Accounting and Economics.* 1979, vol. 1, p. 187 - 210.

Libby, R. 1981. *Accounting and Human Information Processing: Theory and Applications.* Prentice Hall.

Libby, R. 1985. Availability and the Generation of Hypotheses in Analytical Review. *Journal of Accounting Research.* Autumn1985, vol. 23, no. 2, p. 648 - 667.

Loebbecke, J K. and Steinbart, P J. 1987. An Investigation of the use of Preliminary Analytical Review to provide Substantive Audit Evidence. *Auditing: A Journal of Practice and Theory.* Spring 1987, vol. 6, no. 2, p. 74-89.

McDonald, B. and Morris, M., 1984 The Statistical Validity of the Ratio Method in Financial Analysis: An Empirical Examination. *Journal of Business Financial And Accounting*, Spring 1984, vol.11, no. 1, p. 89 - 98.

McDonald, B. and Morris, M., 1985. The Functional Specification of Financial Ratios: An Empirical Examination. *Journal Accounting and Business Research*, Summer 1985, p. 223 – 228.

Mautz, R. K. and Sharaf, H. A. 1961. *Philosophy of Auditing*, Monograph No. 6, American Accounting Association.

Mulligan, C. and Inkster, N. 1999. The Use of Analytical Procedures in the United Kingdom. *International Journal of Auditing,* 1999, vol. 3, no. 2, p. 107 – 120.

Power, M. 1997. *The Audit Society: Rituals of Verification.* Oxford.

Salmi, T, Dahlstedt, R., Luoma, M. and Laakkonen, A. 1988. Financial Ratios as Predictors of Firms' Industry Branch. *The Finnish Journal Of Business Economics,* 1988 vol. 4, p. 263-277.

Statements of Auditing Standards, 1995. *The Auditing Practices Board.*

Statements of Auditing Standards, 1995. SAS410: Analytical Procedures. *The Auditing Practices Board.*

Silverman, B. W. 1985. Some Aspects of the Spine Smoothing Approach to Non-Parametric Regression Curve Fitting. *Journal of the Royal Statistical Society.* 1985 Series B. vol. 47, no. 1, p. 1-52.

Smith, D. G. 1983. *Analytical Review: A Research Study*. Canadian Institute of Chartered Accountants.

Smith, K. A. 1972. The Relationship of Internal Control Evaluation and Audit Sample Size. *The Accounting Review,* 1972, p. 260-269.

Smith, T. M. F. 1979. Statistical Sampling in Auditing: a Statistician's Viewpoint. *The Statistician,* 1979 vol 28, no. 4, p. 276-280.

Steel A. 1992. *Audit Risk and Audit Evidence.* Academic Press.

Stewart, T. 1978. Regression Analysis in Auditing. *Accountancy*, March 1978, p.69-72.

Stringer, K W. 1975. A Statistical Technique for Analytical Review. *Journal of Accounting Research,* Supplement1975, vol. 13, p. 1 - 9.

Tabor, R. H. and Willis, J. 1985. Empirical Evidence on the changing role of Analytical Review. *Auditing: A Journal of Practice and Theory,* Spring 1985, vol. 4, no. 2, p. 93 - 109.

Tukey, J. W. 1977. *Exploratory Data Analysis.* Addison Wesley/Reading. Mass.

Turley and Cooper 1990. *Auditing in the UK.* Prentice-Hall ICAEW.

van Batenburg, P.C., O'Hagan, A. and Veenstra, R.H., 1994. Bayesian discovery sampling in financial auditing: a hierarchical prior model for substantive test sample sizes. *The Statistician,* 1994, vol. 43, number 1, p. 99 –110.

Wallace, W. A. 1983. Analytical Review: Misconceptions, Applications and Experience - Part 1. *The CPA Journal,* January 1983, p. 24 – 37.

Wallace, W. A. 1983a. Analytical Review: Misconceptions, Applications and Experience - Part 2. *The CPA Journal,* February 1983, p. 18 - 27.

Wallace, W. A. 1983b. The Acceptability of Regression Analysis as Evidence in a Courtroom - Implications for the Auditor. *Auditing: A Journal of Practice and Theory,* Spring 1983, p. 66 - 90.

Westwick, C. 1981. Do the Figures Make Sense? *A Practical Guide to Analytical Review.* ICAEW.

Whittington, G. 1980. Some Basic Properties of Accounting Ratios. *Journal of Business Finance and Accounting,* 1980, vol. 7, no. 2, p. 219 – 232.

Wright, A. and Ashton, R. H. 1989. Identifying Audit Adjustments with Attention-Directing Procedures. *The Accounting Review*, October 1989, vol. LXIV, p. 710-728.

York, D. 1997. Assessing Risk, *Accountancy*, ICAEW. February 1989, p. 76.