MODELING AUDITOR JUDGMENT IN NONSTATISTICAL SAMPLING

by

WILLIAM J. READ

dissertation submitted to the Faculty of the

VIRGINIA POLYTECHNIC INSTITUTE and STATE UNIVERSITY

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Business

with a major in Accounting

APPROVED:

ANDREW H. BARNETT   THIRWALL W. BONHAM

DAVID A. BOWNAS   LARRY N. KILLOUGH

WAYNE E. LEININGER   JIMMY W. MARTIN

April, 1984

Blacksburg, Virginia

ACKNOLWEDGEMENTS

I am sincerely grateful to Professors Andrew H. Barnett, David A. Bownas and Wayne E. Leininger for their guidance and assistance throughout this project and to the other members of my committee, Dean Thirwall W. Bonham, Professors Larry N. Killough and Jimmy W. Martin for their insightful critiques.

Bob Tallman, a long-time friend, has been an inspirational force throughout my college career. In the many hours of conversation that we shared he continually stressed the value of education.

My parents, Mary T. and the late William J. Read, were a source of love and support.

The mere presence of my loving wife, Janet, and our two beautiful children, Nathan and Karen, considerably lessened the emotional burden of doctoral study. I shall never forget the sacrifices that they endured so that my educational objectives were obtained.

Finally, this dissertation is dedicated to my grandmother, Mathilda M. Sullivan, who would have celebrated her ninety-fourth birthday on the day (April 9, 1984) it was defended.

TABLE OF CONTENTS

ACKNOLWEDGEMENTS ii

Chapter

page

I. INTRODUCTION 1

Nature of the Problem 1
Identification of the audit population of interest 4
Determine the appropriate sample size 5
Sample Selection 8
Evaluation of the results 9
Consideration of qualitative factors 9
Significance of the problem 10
Statement of the problem and objectives 11
Research Methods and Procedures 12
Organization of the Study 16

II. AUDIT SAMPLING: AN HISTORICAL PERSPECTIVE 18

Evolution of Audit Sampling 18
Auditing Pre-1875 19
Auditing 1875-1940 22
Early Auditing (1900-1929) in the United States 24
Auditing in the United States--1929 to Present 28
Sampling and the Auditor 33
The Objectives of Sampling in Auditing 38
Representative Sampling 38
Corrective Sampling 39
Protective Sampling 40
Preventive Sampling 42
Sampling Methods 43
Nonstatistical (Judgmental) Sampling 43
Statistical Sampling 45
What is Statistical Sampling? 47
Advantages of Statistical Sampling 52
Quantification 53
Objectivity 54
Efficiency 55
Influence of Statistical Sampling on Nonstatistical Sampling 56
Sample Size Estimation 58
Evaluation of Sample Results 59

Chapter Summary 65

III. LITERATURE REVIEW 70

Authoritative Literature 71
Tolerable Error 73
Allowable Risk of Incorrect Acceptance . . . 75
Population Characteristics 78
Lens Studies in Accounting and Auditing 83
Internal Control Evaluation 86
Prediction of Bankruptcy 97
Materiality Judgments 105
Other Selected Studies 112
Chapter Summary 120

IV. THE EMPIRICAL STUDY: RESEARCH METHODOLOGY 123

The Lens Model 124
Indices of Judgment 130
Research Questions 134
Judgmental Accuracy 137
Judgmental Agreement 138
Judgmental Insight 140
The Population 141
Development of the Questionnaire 145
Demographic, Attitudinal, and Professional Characteristics 153
Self-Insight Questionnaire 155
Pilot Study 158
Statistical Analysis Employed 160
Multiple Regression 161
Analysis of Variance 166
Pearson Product Moment Correlation 168
Spearman Coefficient of Rank Correlation . 169
Chapter Summary 170

V. RESULTS OF THE EMPIRICAL STUDY 172

Auditor Characteristics 173
Demographic Characteristics 173
Professional and Behavioral Characteristics 175
Attitudinal Characteristics 179
Results of Statistical Tests 185
Multiple Regression 185
Analysis of Variance 193
Indices of Judgment 203
Judgmental Accuracy 203
Judgment Agreement (Consensus) 208
Agreement in Principle 209

Auditor Self-Insight 210
Chapter Summary 211

VI. SUMMARY AND CONCLUSIONS 214

Summary of the Problem 215
Empirical Findings 217
Conclusions and Recommendations 222
Limitations of the Study 226
Suggestions for Further Research in Audit Sampling 227

BIBLIOGRAPHY 230

Appendix

page

A. MATRIX OF CASES 237

LIST OF TABLES

Table

page

1. Independent Factors' Values 152
2. Educational Characteristics 174
3. Years Associated with Present Firm 176
4. Auditing Experience 178
5. Understanding of SAS No. 39 180
6. Effect of SAS No. 39 182
7. Reaction to SAS No. 39 184
8. Regression Analysis of Group Study 188
9. ANOVA Summary Table for between Auditor Analysis .. 194
10. Duncan's Multiple Range Test 197
11. Auditor R-Square 199
12. Explanatory Factors 200
13. Correlations between Four Factors and Judgments for Eight Auditors 202

LIST OF FIGURES

Figure

page

1. Flowchart of SAS No. 39 Concepts 82

2. Brunswik's Lens Model 128

3. Indices of Judgment 131

4. Case Example 148

5. Demographic, Attitudinal, Professional, and Behavioral Characteristics 154

6. Self-Insight Questionnaire 157

7. Factor Data 192

Chapter I

INTRODUCTION

1.1 NATURE OF THE PROBLEM

Since its issuance in June 1981, Statement on Auditing Standards (SAS) No. 39, "Audit Sampling," has been the center of much controvery. SAS No. 39 defines audit sampling as the application of an audit procedure to less than 100 percent of the items within an account balance or class of transactions for the purpose of evaluating some characteristic of the balance or class.[1] Audit sampling can be statistical or nonstatistical according to the SAS; however, the Statement will have its greater impact on auditors who use a nonstatistical sampling procedure. Although SAS No. 39 contains no specific documentation requirements, the documentation standards set forth in the statements on auditing standards dealing with documentation apply to audit sampling applications just as they apply to other auditing applications. For example, SAS No. 22, "Planning and Supervision", states in part that the auditor should prepare a written audit program[2](THRESH)(OFF)) and SAS No. 41, "Working Papers",

---

[1] Auditing Standards Board, "Audit Sampling," Statement on Auditing Standards No. 39, New York: AICPA, 1981: paragraph 1.

[2] Auditing Standards Board, "Planning and Supervision,"

requires the auditor to prepare working papers that record the work that the auditor has done and the conclusions that he has reached concerning significant matters.[3] Thus, with regard to audit sampling applications, the auditor's audit program might document such items as the objectives of the sampling application and the audit procedures related to those objectives.

The Statement on audit sampling is most controversial for those practitioners who are presently employing a non-statistical sampling application. For example, in using non-statistical sampling to estimate sample size the auditor must be aware of (1) which factors are most important in determining sample size, and (2) the influence of such factors on sample size. SAS No. 39 is quite explicit in identifying and defining those factors that should be considered by the auditor in determining the number of items to be sampled.

The nature of the problem focuses on the concerns of practitioners who anticipate difficulties in designing, selecting, and evaluating a nonstatistical sampling procedure in accordance with SAS No. 39.

Statement on Auditing Standards No. 22, New York: AICPA, 1978.

[3] Auditing Standards Board, "Working Papers," Statement on Auditing Standards No. 41, New York: AICPA, 1982.

Prior to SAS No. 39, most of the larger audit firms used statistical sampling techniques. For these firms, the Statement presents few problems. But, for the practitioner who has relied on nonstatistical sampling where it was previously not necessary to identify the amount of sampling risk, SAS No. 39 has effectively changed the manner in which such auditors sample. With the promulgation of SAS No. 39, factors such as population variation, tolerable error (materiality), expectation of error, and sampling risk must now be considered by the auditor in all sample designs.

SAS No. 39 discusses audit sampling in the context of two factors, namely (1) compliance tests of internal accounting controls, and (2) substantive tests of details. The focus in this study is on nonstatistical sampling in substantive tests of details. The emphasis on substantive tests of details was selected because the audit approach that firms typically use in small business audits is a substantive approach. That is, if the internal accounting control system is not adequately designed to permit reliance, then there is no need or justification in doing compliance testing.

The auditor's judgment process can be classified according to the following five phases of a nonstatistical application:

1.1.1 Identification of the audit population of interest

No matter how the audit is managed or the extent of utilization of others' work, the overall technical objective is to obtain sufficient competent evidence on which to base a report or opinion. The overall objective is obtained by accumulating evidence about several practical audit objectives. These practical objectives are to obtain evidence pertaining to the authenticity of five key assertions in the financial statements. Statement on Auditing Standards No. 31, "Evidential Matter," uses the categories (1) existence or occurence, (2) completeness, (3) rights and obligations, (4) valuation or allowance, and (5) presentation and disclosure.[4] The major point to express about practical audit objectives and their related assertions is that they serve as focal points for organizing procedures. If auditors can discern each objective for each audit problem, then they can more easily identify the sources of competent evidence (i.e., what is the appropriate population) and the procedural means of obtaining the evidence.

---

[4] Auditing Standards Board, "Evidential Matter," Statement on Auditing Standards No. 31, New York: AICPA, 1980: paragraph 15.

1.1.2 Determine the appropriate sample size

As discussed in SAS No. 39, the sample size necessary to provide sufficient competent evidential matter depends on both the objectives and the efficiency of the sample. For a given objective, the efficiency of the sample relates to its design; one sample is more efficient than another if it can achieve the same objectives with a smaller sample size. In general, careful design can produce more efficient samples.[5]

If the auditor plans too small a sample, the sample results will not meet his planned objective. In that situation, the auditor ordinarily needs to perform additional procedures to gather sufficient evidential matter to reach a conclusion. If the auditor plans too large a sample, he examines more items than is necessary to achieve his planned objective. In both cases the examination would be effective even though the auditor did not use sampling efficiently.

There is no rule-of-thumb that is appropriate for all applications. SAS No. 39 imposes no requirement to use quantitative aids, such as sample size tables, to determine sample size. Nor does SAS No. 39 impose a rule regarding minimum sample size. Just as before SAS No. 39, judgment is the key in determining the extent of audit tests.

---

[5] Auditing Standards Board, "Audit Sampling," Statement on Auditing Standards No. 39, New York: AICPA, 1981: paragraph 5.

In determining the appropriate sample size for a substantive test of details, the auditor using nonstatistical sampling must consider the following factors in arriving at his decision:

(a) Variation within the population: The characteristics of individual population items (such as their values) often vary significantly. The auditor should consider this variation when determining the appropriate sample size for a substantive test. The appropriate sample size generally decreases as the variation in the population becomes smaller.[6]

(b) Risk of incorrect acceptance: As discussed in SAS No. 39, an auditor relies on the internal accounting controls, analytical review procedures, and substantive tests of details in whatever combination he believes adequately controls ultimate risk. If the auditor places greater reliance on internal accounting controls, he can accept a greater risk of incorrect acceptance for the planned substantive test. As the acceptable risk of incorrect acceptance increases, the required sample size for the substantive test decreases. Conversely, if the auditor places less reliance on the internal accounting controls, the acceptable level of risk of incorrect acceptance decreases and the appropriate

---

[6] American Institute of Certified Public Accountants, Statistical Sampling Subcommittee, Audit and Accounting Guide: Audit Sampling, New York: AICPA, 1983: p. 55.

sample size increases. The same relationship holds true for the auditor's reliance on the other substantive tests, including analytical review procedures related to the same audit objectives.[7]

(c) Tolerable error and error expectation: The auditor should also consider tolerable error in determining the appropriate sample size for a substantive test. For a given account balance or class of transactions, the sample size required to achieve the auditor's objectives at a given risk of incorrect acceptance increases as the tolerable error for that balance or class decreases.

The auditor considers the amount and frequency of errors that he expects to exist in the account balance or class of transactions when he determines the appropriate sample for a substantive test of details. As the size or frequency of expected errors decreases, the appropriate sample size also decreases. Likewise, as the size or frequency of expected errors increases, the appropriate sample size also increases.[8]

(d) Population Size: The size of the population should have little influence on the auditor's determination of the appropriate nonstatistical sample size for substantive tests.

---

[7] Ibid, p. 56.

[8] Ibid, p.56.

Consequently, it is often inefficient to determine a sample size as a fixed percentage of the population.

1.1.3 Sample Selection

In selecting the sample, the auditor generally identifies individually significant items, stratifies the items subject to sampling, and allocates the sample size to the specific strata.

When the auditor plans the sample, he uses his judgment to determine which items, if any, in the account balance or class of transactions should be individually tested and which items should be subject to sampling. In selecting the sample the auditor separates those individually significant items from the rest of the population. For example, in testing an entity's accounts receivable balance, the auditor might have identified eight customer balances which equalled, or exceeded, tolerable error.

For efficiency, the auditor generally stratifies the remaining population; that is, the items subject to sampling. For example, the accounts receivable balance may include some large dollar balances and many smaller dollar balances. In that case, the auditor might design his sample to include two strata: one for large dollar balances and one of smaller dollar balances.[9]

1.1.4 Evaluation of the results

From the results of examining sampling units, an auditor formulates conclusions, generalizing the results to the population audited. Sample results can be evaluated in terms of the 'achieved sampling risk'. The auditor can then compare this 'achieved sampling risk' with his predetermined level of allowable risk.

1.1.5 Consideration of qualitative factors

In addition to professional judgments about acceptable risk and tolerable error, qualitative elements of analysis include follow-up of all monetary errors discovered in the sample. The auditor should determine whether errors resulted from: (1) irregularities--intentional acts to misstate amounts in the accounts--or (2) errors--inadvertent mistakes affecting monetary amounts. Also, auditors will want to know whether irregularities or errors involve: (a) systematic misstatements or random occurrences, (b) misapplication of an accounting principle, (c) certain types of accounts or transactions (e.g., transactions with related parties), or (d) cases of management override of controls. For example,

---

[9] Ibid, p. 60.

the audit of accounts receivable may indicate the systematic error of overstating customers' balances due, because discounts allowed are recorded sometime after payment is received.

My primary interest in this study is in phase #2 (determining the appropriate sample size) that is, in identifying those significant factors (or cues) which underlie the auditor's judgmental process with respect to the number of items that he elects to sample in a substantive test of details. Of particular concern is the issue of whether auditors are applying the guidelines of SAS No. 39 properly?

1.2 SIGNIFICANCE OF THE PROBLEM

In the course of an audit engagement, auditors must make numerous judgments and express an opinion on the basis of the information and test results that are never perfectly diagnostic with respect to the underlying economic state of the client. The auditor's decision as to sample size in substantive tests of details is but one example of such a judgment. The ability of auditors to appropriately formulate such judgments is crucial, since they may be held liable at common law or under the federal securities laws should the audited financial statements prove to be materially in error. Thus, this threat of litigation accentuates the importance of improving the effectiveness of audits. In addition, appropriate decisions on sample size impact the efficency of audits. As competition among auditors for clients continues to increase, there is an increasing emphasis on increasing the efficiency of audits. The significance of this research is underscored when one considers the importance of audit effectiveness and audit efficiency.

1.3 STATEMENT OF THE PROBLEM AND OBJECTIVES

The premise upon which this research is based is that practitioners are encountering difficulties in designing, selecting, and evaluating a nonstatistical sampling procedure in accordance with the mandate of SAS No. 39. Consequently, the auditor's determination of the appropriate extent of testing is inconsistent with his judgment as to the assurance level he needs from his sampling application, or conversely, the degree of risk that he is willing to accept. The objectives of this study are as follows:

(1) to identify which factors are used by auditors in making nonstatistical sample size judgments in substantive tests of details;

(2) to determine if the nonstatistical sample size judgments of the typical auditor correspond to the 'mean estimation' statistical sample sizes;

(3) to ascertain whether there are statistically significant differences in the nonstatistical sample size judgments of individual auditors;

(4) to measure empirically the auditor's accuracy in making nonstatistical sample size judgments versus the sample sizes that are produced by the 'mean estimation' approach;

(5) to assess the level of consensus across auditors in their nonstatistical sample size judgments;

(6) To determine the amount of 'self-insight' that auditors exhibit into their judgment policies;

(7) to reexamine the need for revision in the authoritative guidelines on 'audit sampling' in the context of the empirical insight developed from this study and the inferences drawn from the literature.

1.4 RESEARCH METHODS AND PROCEDURES

This research can be classified as exploratory research into identifying which factors underlie the auditor's judgment with respect to nonstatistical sample size decisions in substantive tests. As such, this study will utilize Brunswik's Lens Model to provide mathematical representations of the auditor's judgment process.

The research will involve both a group and individual analysis. The group study will involve the random assignment of each of seventy auditors, who will be chosen from the population of practitioners who primarily perform nonstatistical sampling, to one cell in a 7X10 matrix. The matrix will consist of seventy cells, each of which will contain five audit sampling cases for which the subject will be asked to render nonstatistical sample size judgments. A total of fifty audit cases will be constructed for this research. A random permutation of the fifty cases within each of the seven-columns will insure that each case is analyzed once by seven different auditors.

After each auditor has had the opportunity to analyze the five cases that he has been assigned to review, the average recommended sample size of the seven analyses for each of the fifty cases will be determined. To infer which factors are used by the typical auditor in making nonstatistical sample size decisions, the mean recommendations for each of the fifty cases will be regressed onto the independent factors. This resulting multiple linear equation will represent an estimate of the typical auditor's judgment policy relative to nonstatistical sample size decisions.

The subjects selected for this research will be auditors who are representaive of the typical practitioner who primarily audits small businesses. It is from this population that nonstatistical sampling receives its most frequent application.

The group analysis will provide evidence as to how the typical auditor is making nonstatistical sample size decisions in substantive tests of details. That is, it will identify the factors that are used by the typical auditor in nonstatistical sampling applications. Moreover, such empirical findings will provide evidence as to whether the typical auditor properly considers those factors (e.g., population variation, tolerable error, etc.) that are enumerated in SAS No. 39. The advantages of the group analysis in which each practitioner renders sample size judgments on five unique cases are: (1) systematic effects attributable to fatigue and boredom are minimized, and (2) from a practitical perspective, the small case load for each subject permits the selection of a larger sample of practitioners.

Should the group analysis reveal that the typical auditor is properly considering the factors that SAS No. 39 describes, the analysis is inappropriate to assess whether or not significant differences in sample size judgments exist between auditors. Because of this limitation, an individual analysis will also be performed.

The individual analysis will involve a sample of eight auditors who are representative of the typical small practitioner. Each of these eight auditors will be asked to make nonstatistical sample size decisions for each of the fifty cases that will be used in the group analysis. Each auditor's judgment policy can then be inferred by regressing the subject's fifty sample size decisions onto the independent factors. The individual analysis is most appropriate to test the significance of differences in auditor's judgment policies because of the completely crossed factorial design. Thus, the individual analysis will strengthen and complement the group analysis.

Prior to commencing the group and individual analyses, a pilot test will be conducted for the purpose of insuring that the cases are most representative of the small business environment. Any ambiguities in the case narration, the description of the factors or in the required task will be corrected so that each case has maximum external validity.

The pretest will also provide information as to how aversive an analysis of the cases will be on the subjects. That is, the pretest will provide empirical insight into how the cases should best be administered, so that any systematic effects attributable to fatigue and boredom are minimized.

The remaining research objectives deal with the two focal areas of judgment modeling research: (1) the accuracy of judgments, and (2) the agreement (i.e., consensus) among the judgments of subjects. Correlational statistics will be used to assess judgmental accuracy and agreement in principle. The intraclass correlation will be used to evaluate the consensus or consistency in patterns of recommended sample sizes between the auditors' judgments across the fifty cases.

The need for improvements in the authoritative literature will be examined in the context of the results of the empirical study. Based upon inferences drawn from primary and secondary research, recommendations will be made for a revision in SAS No. 39, "Audit Sampling".

1.5 ORGANIZATION OF THE STUDY

The focus in Chapter II will be on the evolution of audit sampling; specifically, the development of both nonstatistical and statistical techniques. Of particular interest will be an evaluation of the influence which statistical sampling methods have had on nonstatistical sampling applications.

Chapter III will review the research that has utilized Egon Brunswik's Lens Model. Emphasis will be placed on human information processing (HIP) research in both the clinical and auditor judgment spectrums.

Chapter IV will disclose in detail the research methods and procedures. The chapter will include a description of the population under study, the method of creation of the research instrument, administration of the final questionnaire, and statistical analyses employed.

Chapter V will contain a presentation and analysis of the results of the empirical investigation.

Chapter VI will summarize the contents of the previous chapters and discuss recommendations for future research.

Chapter II

AUDIT SAMPLING: AN HISTORICAL PERSPECTIVE

2.1 EVOLUTION OF AUDIT SAMPLING

To appreciate more fully audit sampling from a historical perspective, one must be knowledgeable of those forces that have shaped the auditing profession. During the better than four-hundred and fifty years of recognizable existence of auditors, the objectives and techniques of auditors have changed. A review of the history of auditing provides a basis for understanding the changes that have taken place in the attest function.

If one were to dichotomize the history of auditing into two distinct eras, then it might be appropriate to think of auditing in the period before the year 1875, and auditing since that time. The year 1875 is significant because it marked the advent of the Industrial Revolution and the transition from small, owner-operated businesses to large-scale operations and the emergence of hired management.

Brown's review of the history of auditing serves as a basis for analyzing and interpreting the changes that have taken place in audit techniques and objectives.[10] His per-

---

[10] Brown, Gene R, "Changing Audit Objectives and Techniques," The Accounting Review (October 1962), pp. 696-703.

spective on the history of auditing is relied upon in part in this section.

2.1.1 Auditing Pre-1875

Prior to the start of the fourteenth century, the process of auditing was virtually non-existent. Accounting, which had yet to be systematized, was basically centered about governmental and family units. Its primary function involved the use of two scribes who kept independent records of identical transactions to prevent defalcations within the treasuries of ancient rulers.[11] During this time reporting accuracy was a secondary objective of accounting. For example, the physical taking of inventory was for the purpose of ascertaining the accuracy of the accounting records.

With the demise of the Roman Empire, the growth of auditing kept pace with the rise of double-entry bookkeeping in the highly developed civilization of Northern Italy in the fourteenth and fifteenth centuries. Double-entry bookkeeping was used by the merchants of that time in Italy, the Netherlands, and Southern Germany, where the mercantile influence of the Italians was very strong; however, it was not until the close of the fifteenth century that the art of bookkeeping was systematically transcribed by the mathemati-

---

[11] Ibid, p. 696.

cian-monk, Luca Paciolo.[12]

During the fourteenth and fifteenth centuries, there was no industrial activity to speak of only the mercantile activity of buying and selling. To the extent that any semblance of production processes were in operation, they were in the hands of small craftsmen, closely controlled by the rulers of their guilds.[13] The basic purpose of bookkeeping during these centuries was the accountability of bailiff to master.

Throughout the fourteenth and fifteenth centuries, the purpose of auditing was the prevention and detection of fraud. The 'modus operandi' employed by auditors to accomplish this objective was the detailed verification of transactions against supporting documents. Considering the limited size of mercantile activities, a 100 percent verification was hardly impractical and inefficient. Audit sampling or testing as we know it today was nonexistent; as was a system of internal control to assist in the prevention and detection of fraud. Furthermore, a standardized system of accounting that would enhance accuracy in reporting and fraud prevention had yet to be formulated.

---

[12] Solomons, David, "The Historical Development of Costing," Studies in Costing, (1952), p. 3.

[13] Ibid, p. 3.

With the start of the sixteenth century a limited amount of industrial activity had begun. Merchants distributed their raw material stocks to artisians who were typically employed on a piece-rate basis. Accordingly, there was a need for effective methods of controlling stocks of materials in the hands of outworkers, of linking amounts earned by the workers with amounts paid to them, and separation of the financial results achieved by the capitalist on each of his separate activities.[14] The standardized double-entry system of accounting, as developed by Paciolo was promoted to fulfill these needs. Despite this transformation in accounting, its primary purpose remained the detection of fraud within the scope of industrial activity.

Aside from this change in accounting, there was little in the period of 1500 to 1875 that would distinguish audit objectives from the periods of the Roman Empire and the mercantile state of Northern Italy. The primary objective of auditing continued to be the prevention and detection of fraud, with this objective assuming even greater importance with the emergence of the absentee form of ownership. As for audit techniques and procedures, detailed verification of transactions against supporting documents, remained the well-established norm, with the technique of audit sampling

---

[14] Ibid, p. 3.

still an unknown. Moreover, an internal system of checks and balances to assist in the detection and prevention of fraud, had yet to be developed and implemented.

Thus, auditing from the days of the Roman Empire to the birth of the Industrial Revolution, could be characterized by the following:

(1) A stated objective of the detection of fraud.

(2) A 100 percent detailed verification of transactions.

(3) An absence of audit sampling.

(4) A lack of internal accounting controls.

2.1.2 Auditing 1875-1940

With the start of the Industrial Revolution in England in the middle of the nineteenth century, the character of auditing was soon to change. The voluminous business entities that resulted from the Industrial Revolution served as the incentive that projected the corporate form of entity to the forefront. The rampant growth of business operations necessitated the hiring of professional managers to oversee the interests of absentee owners. Prior to the advent of the Industrial Revolution it was not uncommon for an auditor to perform a complete review of the entries and records of the company audited. However, with the industrialization of business enterprises, it became readily apparent that an examination of every transaction in the period covered by the financial statements was inefficient and impractical. Consequently, the auditing profession's only recourse was the implementation of a sampling application, which would provide the auditor with reasonable assurance that the financial statements examined by him are materially correct.

Brown notes that the origin of audit sampling--or test-checking, as it was once called by the auditor--can be traced to a remark made by the presiding judge in the London and General Bank case of 1895:

> Where there is nothing to excite suspicion, very little inquiry will be reasonable and quite sufficient; and, in practice, I believe businessmen select a few cases haphazard, see that they are right and assume that others like them are correct also.[15]

With a change in audit techniques from a detailed verification of transactions to testing, it's plausible to also expect a change in audit objectives. A change in audit objectives usually precipitates a change in techniques, but such was not the situation during this particular period. At the start of the twentieth century, while audit sampling was

---

[15] Brown, Gene R, "Changing Audit Obectives and Techniques," The Accounting Review (October 1962), p. 698.

beginning to be performed, the primary objective of auditing continued to be the detection of fraud.

2.1.3 Early Auditing (1900-1929) in the United States

The objectives and techniques of the auditing profession in the United States in the early 1900's can be traced to the influence of the British auditing profession. Accordingly, the primary objective of auditing in the U.S. during this period was the determination of whether persons in positions of fiscal responsibility were performing and reporting in an truthful manner. Furthermore, the approach to auditing at this time was a balance sheet orientation. The balance sheet approach to auditing considers the income statement accounts as residuals; and holds, that if the balances of the accounts in the balance sheet are properly stated in the aggregate, the results of operations are of necessity properly stated. The popularity of this orientation was because the primary external users of financial data were bankers from whom credit was being requested. As such, bankers had a strong need to assess a firm's debt-paying ability as well as its solvency.

With the growth in size of American business (spurred on by the Industrial Revolution), owners began to employ the services and expertise of professional management. As noted previously, this separation of ownership and management underscored the necessity of auditors to safeguard the interests of owners in absentia. Before the Industrial Revolution, auditing was concerned primarily with the detection and prevention of fraud. However, in the early part of the twentieth century, the focus of auditing started to change. The auditor's determination of whether the financial statements were fairly presented began to supplant the detection of fraud as the primary objective of auditing.

There are at least two important reasons for the shift in auditing emphasis from fraud detection to determining the fairness of the financial statements' representations. First, the ratification of the sixteenth amendment in 1913 gave Congress the power to impose and collect taxes. With passage of the Revenue Act of 1913, the fairness of reported income assumed greater importance.

Second, with the rapid expansion of the securities markets, large numbers of investors demanded an accounting of corporate operating results, earnings per share, and income trends.[16] From a historical perspective, the development of audit sampling is better understood with a knowledge of those forces which influenced the changes in the nature

---

[16] Meigs, Walter B., Whittington, O. Ray, and Robert F. Meigs, Principles of Auditing, (Homewood, Illinois: Richard D. Irwin, 1982), p. 8.

and objectives of auditing.

This significant change in audit objectives was coupled with an important change in techniques. At the turn of the twentieth century, a testing process began to replace the 100 percent verification of entries and transactions.

Despite the increase by auditors in their acceptance of sampling in the early 1900's, there existed a paucity of professional literature on the topic of sampling. One of the earliest pieces was a 'program of audit procedures' which was printed in the Federal Reserve Bulletin in 1917. During this period it was widely believed that the auditor's examination certified the accuracy of the financial statements. The primary purpose of the bulletin was to correct these misunderstandings. But also included in the bulletin were some references to sampling, such as selecting a "few book items" of inventory.[17] The AICPA's earliest precursor, the American Association of Accountants, was responsible for the publication of the 'program of audit procedures.'

Early audit sampling was marked by the failure on the part of the auditor to recognize the need to evaluate the system of internal control as a guide to the direction and amount of sampling to be performed. That is, auditors did

---

[17] American Institute of Certified Public Accountants, Statistical Sampling Subcommittee, Audit and Accounting Guide: Audit Sampling, New York: AICPA, 1982: p. 5.

not discern a relationship between the extent of sampling and the effectiveness of an entity's system of internal accounting control. Internal control was recognized as a subsystem within the standardized structure of accounting, but little attention was paid to it in auditing.

As auditors became comfortable with and gained insight into the technique of sampling, they became aware of the influence which the system of internal control has on the extent of audit sampling. Some auditing articles and textbooks in the early 1900's referred to reducing the extent of tests of details based on reliance on the entity's 'internal check', as internal accounting control was first called. Brown notes that the British textbook, Audit Programmes, published in 1910, emphasized that the initial step in an audit was to "ascertain the system of internal check".[18] The first article to be published in an American professional accounting journal on the topic of internal control appeared in October, 1913. The author, F. R. Carnegie Steele, demonstrated keen insight with his statement that, "Systems of operating accounts and records should be formulated with a single guiding principle--that they must furnish means of control".[19]

---

[18] Brown, Gene R, "Changing Audit Objectives and Techniques," The Accounting Review (October 1962), p. 700.

[19] Carnegie Steele, F. R., "The Development of Systems of

Despite these and other references in the literature on the importance of the internal check and its association with the amount of sampling; audit practice was slow to take heed. While sampling continued to grow in popularity, the profession as a whole failed to assimilate the linkage between the appraisal of internal controls and the extent of testing. Thus, audit procedures changed, the stated audit objectives changed; but the attitude of the auditor was slow to change.[20]

2.1.4 Auditing in the United States--1929 to Present

It wasn't until 1929 that the American Institute of Accountants formerly recognized the importance of ascertaining the effectiveness of the internal check. Cochrane reports that a bulletin, entitled "Verification of Financial Statements," was prepared by a committee of the Institute, for issuance to auditors, accountants and other concerned parties.[21]

The preface of this bulletin stated that

Control," Journal of Accountancy (October, 1913), p. 282.

[20] Brown, Gene R., "Changing Audit Objectives and Techniques," The Accounting Review (October 1962), p. 700.

[21] Cochrane, George, "The Auditor's Report: Its Evolution in the USA," The Accountant (November 1950), pp. 448-460.

> The scope of the work indicated in these instructions includes a verification of the assets and liabilities of a business enterprise at a given date; a verification of the profit and loss account for the period under review and, incidentally, an examination of the accounting system for the purpose of ascertaining the effectiveness of the internal check (emphasis added)....The extent of the verification will be determined by the conditions in each concern.

With the Institute publicly acknowledging that the degree of testing is largely dependent on the effectiveness of internal control; the technique of audit sampling quickly became the rule, not the exception.

In the face of the growimg harmonization within the profession on the linkage between the effectiveness of internal control and the extent of testing; there remained widespread disagreement in the early 1930's as to what should be the audit's primary purpose. While the movement in support of the objective of determining the fairness of financial statement representations was well along; many within the profession remained steadfast to their belief, that the primary objective of an audit was the detection of fraud.

There were a series of critical events that caused the pendulum to swing away from the objective of fraud detection, to the objective of the attest function.

The 1931 court case of Ultramares Corporation v. Touche, emphasized the legal responsibility of the auditor to third parties who can be foreseen by the auditor.[22] In this case, Ultramares, a factor, sued the audit firm for gross negligence in the performance of their audit because of losses they incurred on loans made to the auditor's client. The Ultramares decision is significant because it established a basis for legal action by third-parties against the auditors for gross negligence in the performance of their professional responsibilities. Gross negligence exists when the auditor has rendered an opinion that is not supported by sufficient, competent evidential matter. The case led to a greater emphasis being placed on the development of external support for such things as accounts receivable through confirmation procedures.[23]

The Securities Act of 1933 also played a vital role in the profession's movement toward the primacy of the attest function objective. The Act stated that auditors who express an unqualified opinion in a registration statement concerning the public offering of securities, may be held liable to third parties if the registration includes material errors or omissions. Shortly after passage of the Statute, the Securities and Exchange Commission was created by Congress to

---

[22] Ultramares Corporation v. Touche, (255 N.Y. 170, 174 N.E. 441, 1931).

[23] Thomas, W. C., and E. O. Henke, Auditing: Theory and Practice, (Boston, Kent Publishing Co., 1983), p. 44.

enforce the Act's provisions.

If there is any one particular event that served to have the greatest impact on the auditing profession in recognizing the importance of the attest function, it would be the McKesson & Robbins fraud case. While this dispute was settled out of court, SEC investigators and the AICPA used the facts to stimulate development of professional auditing standards.

Prior to McKesson & Robbins Inc., auditors had denied responsibility for confirming accounts receivable and for verifying the physical existence of inventory quantities. These practices were not standardized procedures in auditing at the time.

In McKesson & Robbins, some company offical s who understood auditing techniques deliberately overstated total assets by approximately $19 million. Out of a total asset base of $87 million, inventory was inflated by approximately $10 million and accounts receivable consisted of $9 million in fictitious accounts.

The McKesson & Robbins case emphasized the need for objectively developed financial statements for which selected account balances are verified against external evidence.[24] Following McKesson & Robbins, the AICPA issued Statements on

---

[24] Ibid, p. 44.

Auditing Standards Numbers 1 and 2, which required auditors to follow more stringent standards in confirming the existence of accounts receivable and merchandise inventory. This landmark case also provided the incentive for the Institute's establishment of the ten generally accepted auditing standards (GAAS), which serve as the cornerstone of audit practice as we know it today.

McKesson & Robbins also prompted the SEC to issue Accounting Series Release (ASR) No. 19, in 1940. In this particular pronouncement, the SEC states that:

> Even though the auditors are not guarantors and should not be responsible for detecting all fraud (emphasis added), the discovery of gross overstatements in the accounts is a major purpose in an audit even though every minor defalcation might not be disclosed.

Following this series of events, there was a fair amount of agreement within the auditing profession that the auditor could not, and should not, be primarily concerned with the detection of fraud.

Thus, the decade of the 1930's was a transitional period for the auditing profession. Audits were no longer directed primarily toward discovering bookkeeping irregularities and fradulent activities on the part of employees. The primary emphasis was now concerned with the auditor's determination of the fairness of financial statement representations.

While some applications of audit sampling can be traced to the period of the Industrial Revolution; it was not until the 1930's that audit sampling became the rule not the exception. Furthermore, the linkage between the extent of testing and the effectiveness of internal control became well accepted and established during this decade.

Since 1940, the role of audit sampling in the gathering of sufficient, competent evidential matter has continued to flourish. From purely judgmental sampling applications to the sophisticated statistical techniques of today, the evolution of audit sampling as an audit procedure has been rapid.

2.2 SAMPLING AND THE AUDITOR

Statement on Auditing Standards Number 1, defines the auditor's objective by stating, "The objective of the ordinary examination of financial statements by the independent auditor is the expression of an opinion on the fairness with which they present financial position and results of operations."[25] Thus, the auditor's overall objective in carrying out the attest function is to make reasonably certain that financial statements examined by him are fairly presented in

[25] Committee on Auditing Procedure, American Institute of Certified Public Accountants, "Codification of Auditing Standards and Procedures," Statement on Auditing Standards No. 1, New York: AICPA, 1973: p. 1.

conformity with generally accepted accounting principles. The auditor therefore wishes to find material errors in the financial statements if they exist. To realize this objective the auditor employs numerous procedures, not the least of which is his sampling application.

To express his opinion on the fairness of the financial statements' representations, the independent auditor studies and evaluates the existing internal control system as a basis for determining the extent of testing. It is insufficient for the auditor to simply assure himself that a satisfactory internal control system is in operation. There still remains the basic concern that although such a system exists, it may exist only in a flowchart.

To gain confidence that the system is in fact operating as prescribed, the auditor examines a portion of the records and entries that constitute the evidence of the effectiveness of the control system.[26] When the auditor examines a portion of the documentation for the purpose of assuring himself that, in fact, the internal control system is effectively operative, he is performing tests of compliance.

The authoritative support for the role of compliance testing can be found in Statement on Auditing Procedures No. 33, which states that "the well-established practice of the

---

[26] Arkin, H., Handbook of Sampling for Auditing and Accounting, (New York: McGraw-Hill, 1974), p. 3

independent auditor of evaluating the adequacy and effectiveness of the system of internal control by testing the accounting records and related data ...has generally proved sufficient for making an adequate examination.[27]

In addition to the deviations from prescribed internal controls that are uncovered through tests of compliance, the auditor also wishes to substantiate the fairness of recorded amounts in the various accounts. The purpose of these substantive tests is the discovery of significant errors in the financial statements. The degree of reliance that the auditor is able to place on internal accounting controls determines the nature, timing, and extent of the substantive testing.

When the auditor makes generalizations regarding the entire population based on sample findings, a test operation has been performed. Sampling takes place anytime generalizations are made from test results about an entire field of data. Thus, sampling is the process of selecting a sample from a larger group of items (called the field or population) and using the characteristics of the sample to draw inferences about the characteristics of the entire field of items. If the auditor is interested only in discussing or

---

[27] American Institute of Certified Public Accountants, Auditing Standards and Procedures, Statements on Auditing Procedures No. 33, New York: AICPA, 1963: p. 11

observing any instances that he may accidentally encounter in a haphazard examination of some entries, without considering their impact on the books of record in terms of the frequency of their occurrence or their magnitude, he has no sampling application.[28]

The sample results are but one piece of evidence that the auditor considers in arriving at an overall opinion regarding the fairness of the financial statements. In other fields; however, the sampler typically places greater reliance on the sample results. For example, the sampler in political surveys typically will arrive at his decision regarding the population of interest on the sole basis of his sample's results. Politicians and their supporters are immensely interested in knowing their prospects of winning an election as the campaign heads toward final balloting. By sampling 1,000 registered voters prior to the election, the percentage who claim they will vote for a given candidate may be used to estimate the percentage of votes the candidate will receive in the election. The estimated percentage could be used to decide, for example, whether a greater campaign effort (more money) is required to assure the candidate's election.

---

[28] Arkin, Herbert, Handbook of Sampling for Auditing and Accounting (McGraw-Hill, Inc., 1963): p. 4.

As noted previously, the role of the independent auditor is to perform an examination permitting him to express an opinion as to the reasonableness of the financial statements and their compliance with generally accepted accounting principles.

To this end, the auditor examines a variety of accounting evidence in accordance with the third standard of field work which states:

> Sufficient competent evidential matter is to be obtained through inspection, observation, inquiries, and confirmations to afford a reasonable basis for an opinion regarding the financial statements under 'examination'.[29]

The need for sufficient, competent evidential matter is paramount since it is the basis for the auditor's expression of opinion. The auditor's dependence on sampling procedures as a means of gathering such evidence is one of the basic reasons the auditor's report is regarded as an expression of opinion, rather than absolute certification of the fairness of financial statements.

---

[29] Committee on Auditing Procedure, American Institute of Certified Public Accountants, "Codification of Auditing Standards and Procedures," Statement on Auditing Standards No. 1, New York: AICPA, 1973: Section 330:01.

2.3 THE OBJECTIVES OF SAMPLING IN AUDITING

Before discussing further the evolution of audit sampling and, in particular, the emergence of statistical sampling techniques, it is appropriate to enumerate the objectives of sampling in auditing.

Regardless of whether the auditor uses a statistical or nonstatistical sampling application, the sampling process in auditing can be characterized by four objectives. These objectives are labeled according to the four types of sampling: (1) representative-- sample characteristic equals the population characteristic, (2) corrective-- sample contains the maximum number of dollar value of error items, (3) protective-- sample includes the maximum total dollar value, and (4) preventive-- sample creates the maximum degree of uncertainty about the scope of future audits.[30]

2.3.1 Representative Sampling

Of the four objectives, perhaps the most traditional objective is the one of estimation and acceptance sampling, which can be referred to as representative sampling. The purpose of representaive sampling is to project a characteristic of the audit field, such as the number of items that

[30] Ijiri, Yuji and Robert S. Kaplan, "The Four Objectives of Sampling in Auditing: Representative, Corrective, Protective and Preventive," Management Accounting (December, 1970), pp. 41-44.

are in error in an account.

For example, suppose an accounts receivable subsystem contains 1,000 receivables with a mean value of $100 per account. Of the population of accounts, assume that 50 are recorded in error. If the auditor is interested in estimating the total value of the accounts receivable, the sample would be representative if its mean dollar is $100 per item, while if the auditor is interested in estimating the number of items that are in error, the ideal sample is one in which 5 percent of the sampled items are incorrect.

For the most part, the development of sampling theory in accounting has been constrained by the narrow perspective that the auditor's sampling objectives are focused on representativeness. It is the position of Ijiri and Kaplan that auditors have at least three additional objectives. They further contend that the ideal sample should integrate all four objectives.

2.3.2 Corrective Sampling

As the name denotes, the objective of corrective sampling is to identify and correct as many error items as possible. Corrective sampling is typically applied in those circumstances when the auditor suspects that errors or defalcations have occurred.

One method by which the auditor can apply corrective sampling is through population stratification. The auditor typically stratifies the population of interest so that he can take samples from those strata where errors are more prone to occur. From this perspective, the purpose of corrective sampling is to maximize the collection of error items.

An experienced practitioner should be able to extract a sample which contains significantly more error items than a sample drawn at random from the entire population. Ijiri and Kaplan indicate that it is quite common for good auditors to check heavily the transactions in the end or the beginning of the period, the items that have been recently added or dropped, the accounts with unusually high or low values, or those with negative balances.[31]

2.3.3 Protective Sampling

The objective of protective sampling is to include in the sample the greatest amount of population dollar value. That is, protective sampling is concerned with the auditor obtaining a sample that maximizes the dollar-value of the items included therein.

---

[31] Ibid, p. 43.

The objective of protective sampling can be best understood from the aspect of representative and corrective sampling. In implementing representative sampling, if the high-value items are more prone to error, then the auditor's ideal sample should contain a disproportionate number of such items. Similarly, in implementing corrective sampling, if the chance of an item being erroneously recorded for high-value items exceeds that of low-value items, the auditor seeks to maximize the number of high-value items included in the sample. However, Ijiri and Kaplan note that auditors appear to have an intrinsic attraction to high-value items, since they often include such items in their sample even when such items might be less likely to be in error because of added internal checks.[32] When such is the case, protective sampling with its emphasis on maximizing the proportion of dollar values sampled, can bring the auditor a false sense of security.

It is interesting to compare the contrast in emphasis between protective and corrective sampling. As noted, the objective of the former is to enhance the auditor's assurance by maximizing the proportion of dollar-value items included in the sample. Ijiri and Kaplan consider protective sampling to be a defensive approach since it tends to bring

---

[32] Ibid, p. 43.

the auditor a feeling of greater assurance that the population is free from frauds and errors when he has tested 70 percent as opposed to 30 percent of the total value of the account. Conversely, with corrective sampling the practitioner assumes an offensive stance as he attempts to sample those items which project the greatest liklihood of being in error.

2.3.4 Preventive Sampling

With preventive sampling the practitioner attempts to randomly sample every segment of the population in order to convey the message to the auditee, that no one particular area is above being audited. In pursuit of this objective, the auditor endeavors to not establish a pattern that the client can associate with his sampling application. Thus, preventive sampling is a method of deterring the occurrence of defalcations.

It's worth noting the purpose behind random sampling in both the representative and preventive applications. In representative sampling, the auditor randomly selects so that he can validly make mathematically supportable statements regarding the precision and reliability of the estimate.[33] In implementing protective sampling, the auditor attempts to

---

[33] Ibid, p. 44.

convey to the auditee the idea that no one particular area is free from the scrutiny of the auditor. As such, the goal of preventive sampling is the deterrence of fraud.

With respect to the technique of sampling, the ideal sample should integrate all four objectives. For too long now, it seems that the auditing profession has tended to emphasize representative sampling to the virtual exclusion of the other three objectives.

2.4 SAMPLING METHODS

2.4.1 Nonstatistical (Judgmental) Sampling

Pure judgmental sampling is based strictly on the auditor's professional expertise and professional judgment. The term "judgment" is not meant to connote that the auditor exercises his judgment only in nonstatistical applications; instead, it means that the sample is not supported by the statistical laws of probability.

Although a judgmentally selected sample provides the practitioner with no scientific basis for measuring the risk of material sampling error; the use of judgmental methods is widespread even by practitioners who are proponents of statistical sampling. This is because there are segments of most audits where statistical sampling techniques, despite their inherent objectivity, are inappropriate, inadequate, or simply not cost efficient. For example, nonstatistical samples are applicable when the audit area of interest consists of either a few items of material value or many items whose total value is immaterial. For example, nonstatistical sampling would be used to vouch ten additions to fixed assets worth $50,000, when total additions to the account consist of twenty-five items aggregating $75,000.

Perhaps the most serious threat identified with the use of judgmental sampling is the inability of the auditor to minimize frauds by not taking random samples in all possible control areas. This potential deficiency was discussed earlier within the context of preventive sampling. Specifically, because of each auditor's unique idiosyncratic judgmental process, professional training, cultural biases, etc., there is a tendency to consistently apply the same selection process each time judgmental sampling is performed. When this is the case, the likelihood of the auditee being able to predict which items are susceptible to being sampled, as well as the items that are likely to remain audit-free is increased. With the client potentially being able to prognosticate the auditor's testing process, the danger of frauds occurring and going undetected is increased.

2.4.2 Statistical Sampling

Prior to the 1950's, the sampling that was done in auditing was for the most part of a judgmental nature. It was not until the decade of the 1950's that the spark of interest in statistical sampling was ignited. Perhaps this surge in interest was initiated by the AICPA's Special Report of Committee on Auditing Procedures in 1954. The following is an excerpt from the report:

> "....Tests made haphazardly (judgmentally) are without significance and will be of little comfort to the auditor who is called upon to demonstrate that he has exercised due care in his examination (emphasis added). The objective of testing is to determine whether reliance may be placed upon the examinee's representations as expressed in the books of account and financial statements. The appropriate degree of testing will be that which may reasonably be relied upon to bring to light errors in about the same proportion as would exist in the whole of the record being tested.[34]

While not espousing the adoption of a statistical sampling application per se, the Report nevertheless states, that a judgmentally selected sample could not be objectively defended by the auditor. The Report concludes that the selection of the testing method as well as the interpretation of the test results are largely a matter of the auditor's professional judgment.

---

[34] American Institute of Certified Public Accountants, "Generally Accepted Auditing Standards-Their Significance and Scope," Special Report of Committee on Auditing Procedures (New York: AICPA, 1954), p. 37.

A somewhat more definitive position on statistical sampling was taken by the Committee on Statistical Sampling of the AICPA in 1962. In a special report entitled, "Statistical Sampling and the Independent Auditor", the Committee concluded that statistical sampling methods are useful in some circumstances and are permitted under generally accepted auditing procedures.[35]

Despite its formal sanctioning of statistical sampling, the Committee was quick to point out, that the choice of sampling technique was still a matter of the practitioner's judgment. A year later, the AICPA in its Statements on Auditing Procedures (SAP) No. 33, voiced stronger support for the auditor's adoption of a statistical sampling orientation.

The following excerpt is from SAP No. 33. ..."In determining the extent of a particular audit test and the method of selecting items to be examined, the auditor might consider using statistical sampling techniques which have been found to be advantageous in certain circumstances."[36]

---

[35] American Institute of Certified Public Accountants, "Statistical Sampling and the Independent Auditor," Special Report by the Committee on Statistical Sampling (New York: AICPA, 1962).

[36] American Institute of Certified Public Accountants, Auditing Standards and Procedures, Statements on Auditing Procedures No. 33, (New York: AICPA, 1963), p. 37.

2.5 WHAT IS STATISTICAL SAMPLING?

While judgmental sampling is based solely on the auditor's professional judgment, statistical sampling is rooted in the mathematically proven laws of probability. As such, statistical sampling helps the auditor to (1) design an efficient sample, (2) measure the sufficiency of the evidence obtained, and (3) evaluate the sample results.[37]

By relying on the laws of probability, statistical sampling provides for the quantification and control of material sampling error (risk). Sampling error refers to the risk that is associated with examining only a portion of the population items. It is the risk that the selected sample does not mirror the characteristics of the population.

With respect to substantive tests of details, SAS No. 39 enumerates two types of material sampling risk: (1) the risk of incorrect acceptance--is the risk of the auditor concluding that the recorded book value is not materially misstated when in fact it is, and (2) the risk of incorrect rejection--is the risk of the auditor concluding that the book value as recorded is materially in error when in fact it is not. Statisticians refer to the risk of incorrect acceptance as the risk of a Type II error or beta risk. Similarly, the risk of incorrect rejection is sometimes referred

---

[37] Statistical Sampling Subcommittee, <u>Audit and Accounting Guide</u>: <u>Audit Sampling</u>, (New York: AICPA 1982) p. 13.

to as the risk of a Type I error or alpha risk. Of the two risk classifications, the beta risk which impacts audit effectiveness is perhaps the most serious, for it can cause an unadjusted material error in an account to be recorded in the financial statements. Subsequently, if third parties rely upon such erroneous data and suffer losses, they may have legal recourse against the auditor.

In contrast, alpha risk relates to the cost and efficiency of the audit. That is, should the auditor reject the propriety of the recorded book balance, he will typically employ additional testing and procedures in an attempt to refute his initial findings. Thus, the audit would be inefficient; however, it would not be ineffective.

Conversely, non-sampling error (or risk) cannot be mathematically measured. Non-sampling error refers to the risk that results from (1) human mistakes, such as failing to recognize errors in documents in a sample, (2) applying audit procedures that are inappropriate given the audit objective, and (3) misinterpreting the results of the sample.[38] Although non-sampling risk cannot be quantified, it can be minimized by proper planning and supervision, and by encouraging quality control standards on a firm-wide basis.

---

[38] Walter G. Kell and Richard E. Ziegler, Modern Auditing, (New York: John Wiley & Sons, 1983) p. 181.

A common misconception among auditors is that the terms statistical sampling and random sampling are synomous. While the term random sampling is a component of the statistical sampling process, the two terms are not interchangeable. Random sampling refers to the random selection of items for inclusion in the sample. It is but one step in the entire sampling process.

Thus, statistical sampling is a pervasive concept that encompasses three phases: (1) the determination of the appropriate sample size, given the test objectives, (2) the selection of the sample, and (3) the evaluation and interpretation of the findings. It is this complete three-stage process that is properly labeled statistical sampling.

As noted previously, random sampling is the procedure of selecting the items that the auditor will sample. As such, the process of random selection is applicable to the second phase of the entire statistical sampling process. The importance of random-based selection methods is in the auditor's control of material sampling risk. If auditors are to control the risk of sampling error, random sampling procedures must be used in conjunction with statistical measurement procedures for determining sample size and interpreting sample results.[39]

---

[39] Meigs, W. B.; Whittington, O. R. and R. F. Meigs, Principles of Auditing, p. 235.

The importance of random-based selection methods to the sampling process is underscored by SAS No. 39, paragraph 24:

> Sample items should be selected in such a way that the sample can be expected to be representative of the population. Therefore, all items in the population should have an opportunity to be selected. For example, random-based selection which includes for example, random sampling, stratified random sampling, sampling with probability proportional to size, and systematic sampling with one or more random starts.[40]

Thus, the idea behind random selection is that every item in the population has a chance of being selected in the sample. As such, the selection process is not impaired by the personal bias of the sampler. With the subjective bias of the sampler removed, a sample that has been randomly selected can be properly characterized as an unbiased sample. The auditor's selection of a random sample coupled with the laws of probability enables the risk associated with selecting a sample, which is not representative of the population, to be measured and controlled.

The auditor would be likely to apply a statistical sampling technique when the audit area of interest is composed of a large number of similar items. For example, an auditor would use a statistical application to estimate the percentage of deviation from an established internal control

---

[40] Audit Standards Board, "Audit Sampling," Statement on Auditing Standards No. 39, (New York: AICPA, 1981), paragraph 24.

procedure relating to sales, when 75,000 sales transactions have been processed during the year.[41] Similarly, statistical sampling would be used by the auditor to determine whether a $500,000 book value of trade accounts receivable is a materially accurate measurement of the "actual" current accounts receivable total. As a general rule, statistical sampling methods should be used when (1) selection of an efficient random sample is feasible, (2) the mathematical assumptions underlying quantitative statistical calculations are apparently met, and (3) the audit team wants to obtain an objective quantitative measure of the evidential results.[42]

While the advantages identified with a statistical sampling application are quite persuasive, SAS No. 39 specifically states that ..."either a statistical or nonstatistical approach to audit sampling, when properly applied, can provide sufficient evidential matter in accordance with the third standard of field work."[43]

---

[41] Thomas, C. W. and E. O. Henke, Auditing: Theory and Practice (Boston: Kent Publishing, 1983), p. 301.

[42] Robertson, J. C. and F. G. Davis, Auditing (Plano, Texas: Business Publications, Inc., 1982) p. 361.

[43] Audit Standards Board, "Audit Sampling," Statement on Auditing Standard No. 39, (New York: AICPA, 1981), paragraph 04.

For the auditor his choice of statistical or nonstatistical sampling, does not eliminate the need for professional judgment. For example, the auditor's sampling application does not directly influence how he evaluates the competence of the evidence that is procured.

2.6 ADVANTAGES OF STATISTICAL SAMPLING

Because of conceptual and economic reasons, the auditor is faced with rendering an opinion as to the fairness of the financial statements' assertions even though he has examined less than 100 percent of the underlying data. The auditor is justified in accepting some uncertainty that is associated with audit sampling, when the cost and time required to make a 100 percent examination of the data are, in his judgment, greater than the adverse consequences of possibly expressing an erroneous opinion from examining only a sample.[44] Furthermore, the third standard of field work, which deals with evidential matter, requires that the auditor only have a 'reasonable', not absolute, basis for his opinion. Thus, in the conduct of his audit, this effectively means that the auditor may err when he relies on an internal accounting control, or when he concludes that a recorded book value is not materially misstated. Consequently, financial statements

---

[44] Kell, Walter G. and R. Ziegler, Modern Auditing, p. 180.

that may be materially in error, may be issued as being 'fairly presented' because the evidence is simply not sufficient to enable the auditor to discover that the control is nonexistent in practice, and/or that an account is unfairly stated.[45] As noted previously, the risk that is identified with examing only a portion of the population items is known as sampling risk. It is in the measurement and control of material sampling risk that statistical sampling techniques can benefit the auditor. The advantages of statistical sampling will be discussed within the context of three features: (1) Quantification, (2) Objectivity, and (3) Efficiency.

2.6.1 Quantification

The greatest advantage in using statistical sampling is that the risk of material sampling error may be quantified and controlled. That is, statistical sampling provides the auditor with specific statistical measures, such as precision and reliability, that will enable him to control the extent of his risk in relying on sample results. Thus, the auditor is able to specify prior to selecting his sample, the reliability or confidence he requires in his sample re-

---

[45] Bailey, Andrew D., Jr., Statistical Auditing: Review, Concepts and Problems (New York: Harcourt Brace Jovanovich, 1981), p. 3.

sults. Precision might be used in audit sampling as an evaluation concept. SAS No. 39 in this context uses the concept of an 'allowance for sampling risk.'

SAS No. 1 defines precision as "the range of values, plus or minus, around the sample result, and reliability or confidence, which is expressed as the proportion of such ranges from all possible similar samples of the same size that would include the actual population value."[46] the specified precision interval. The complement of the reliability percentage, the alpha risk, represents the risk that, as a result of the auditor taking a sample, his sample estimate may differ from the true population value by more than pre-established precision limits.

2.6.2 Objectivity

Objectivity is the second most important advantage of statistical sampling. In the context of statistical sampling, objectivity relates to the random selection of items for inclusion in the sample. The random selection process provides for disclosing conditions or factors which might not be considered or even known of in choosing a sample by judgment or intuition to say nothing of just plain hapha-

[46] Committee on Auditing Procedure, American Institute of Certified Public Accountants, "Codification of Auditing Standards and Procedures," Statement on Auditing Standards No. 1, (New York: AICPA: 1972), Section 320.03A.

zardly.[47] The importance of objectivity is in the fact that if auditors are to control material sampling error, random sampling procedures must be used in conjunction with statistical measurements techniques for determining sample size and interpreting sample results.

2.6.3 Efficiency

Statistical sampling is applied when the population of interest is too large, complex, etc., to evaluate given the auditor's objectives. Because sample sizes necessary for specified confidence levels and precision estimates are not a direct function of population size, economies of scale can be realized if the auditor employs a statistical sampling technique when the population data is quite voluminous. As noted previously, not only are statistical sampling plans designed to provide sample sizes of sufficient size to produce estimates of specified precision and reliability; but, such plans produce sample sizes that enhance audit efficiency.

During the administration of this study's experiment, many auditors expressed the opinion that a fixed, constant percentage of items must be tested regardless of the population total. While this type of nonstatistical sampling may

---

[47] Hill, Henry; Roth, Joseph, and Herbert Arkin, Sampling in Auditing, (New York: Ronald Press, Inc., 1962), p.7.

be audit effective, it most assuredly leads to audit inefficiency and the waste of time and resources.

In summary, both statistical and nonstatistical sampling can provide the auditor with sufficient, competent evidential matter in accordance with the third standard of field work. Moreover, both types of sampling applications are subject to sampling and nonsampling error. The primary advantage associated with statistical sampling is the quantification and control of material sampling error. Furthermore, statistical sampling benefits the auditor by providing him with an objective basis to evaluate the evidence he has gathered. Statistical sampling does not however, reduce the need for auditors to make judgments. For example, the auditor must exercise judgment in his design of sampling applications and in the inferences that he draws from his test results.

2.7 INFLUENCE OF STATISTICAL SAMPLING ON NONSTATISTICAL SAMPLING

The role of sampling has continued to increase in the development of contemporary auditing. Given the size and complexity of today's audit client, the demand for statistical sampling techniques has also continued to increase. Although the benefits identified with statistical sampling are well documented, the auditor remains free to elect either a statistical or nonstatistical sampling application. This is not meant to infer that nonstatistical sampling has not been influenced to a considerable degree by statistical sampling concepts. One need not look any further than SAS No. 39 to understand and appreciate the influence that statistical sampling has had on nonstatistical sampling.

As noted previously, SAS No. 39 is applicable to both statistical and nonstatistical sampling. The Statement also applies to both compliance and substantive testing. Because this study is primarily concerned with nonstatistical sample size decisions in substantive tests of details, the influence of statistical sampling concepts on nonstatistical sampling will be discussed within this context.

The primary differences between statistical and nonstatistical sampling are in the steps for ascertaining sample size and in evaluating sample results.[48] These steps are perceived as being more rigorous and objective in statistical sampling and more judgmental and subjective in nonstatistical sampling.[49]

---

[48] Kell & Ziegler, Modern Auditing, p. 377.

[49] Ibid, p. 377.

2.7.1 Sample Size Estimation

In using nonstatistical sampling to estimate sample size, the auditor must be aware of (1) which factors are most important in determining a sample size, and (2) the influence of such factors on sample size. SAS No. 39 is quite explicit in identifying those factors that should be considered by the auditor in his determination of the number of items to be sampled. Although all the factors are highly judgmental, the auditor's working papers should verify that such factors were in fact considered. With the promulgation of SAS No. 39, factors such as (1) population size, (2) population variation, (3) tolerable error, (4) expected error, (5) risk of incorrect acceptance, and (6) risk of incorrect rejection, must now be implicitly considered by the auditor in his nonstatistical sample size decisions. Careful subjective analysis of these factors in a particular circumstance, coupled with the auditor's seasoned experience and professional judgment, should result in a sample size that is more appropriate than an arbitrarily determined sample.[50]

In designing nonstatistical sample sizes in substantive tests of details, it is important that the auditor understand the relationship between each of the factors and sample size. That is, the auditor should implicitly consider in

---

[50] Ibid, p. 378.

his judgment process each of the following relationships:[51]

FACTOR	EFFECT ON SAMPLE-SIZE
Risk of incorrect acceptance	Inverse effect
Risk of incorrect rejection	Inverse effect
Population size	Small direct
Population variation	Direct effect
Tolerable error	Inverse effect
Expected error	Direct effect

When the auditor uses statistical sampling, he explicitly considers the above factors by their inclusion in mathematical formulas. As a result, statistical sample sizes are both audit effective and audit efficient.

2.7.2 Evaluation of Sample Results

When the auditor uses a nonstatistical sampling application he is not able to objectively evaluate his results in terms of achieved precision. However, drawing on the con-

---

[51] Ibid, p. 378.

cepts of statistical sampling, the auditor may subjectively compare his pre-determined tolerable error with the error he projects as a result of his test findings. According to SAS No. 39, the total projected error for a sample "should be compared with the tolerable error for the account balance or class of transactions, and appropriate consideration should be given to sampling risk."[52] If the total projected error is less than tolerable error for the account balance or class of transactions, the auditor should consider the risk that such a result might be obtained even though the true monetary error for the population exceeds tolerable error.[53] Conversely, if projected error approximates or exceeds tolerable error, the auditor may be justified in concluding that there is an intolerable sampling risk that the tolerable error is considerably less than the true error. In such a case, the auditor performs additional testing to determine the true value of the account with greater precision and confidence levels. However, unlike the objective evaluation afforded the auditor who uses statistical sampling, the auditor's evaluation of results in a nonstatistical context are purely subjective and thus, less defensible.

---

[52] Auditing Standards Board, "Audit Sampling," _Statements on Auditing Standards No. 39_, (New York: AICPA, 1981), paragraph 26.

[53] _Ibid_, paragraph 26.

Prior to SAS No. 39, most of the larger audit firms used statistical sampling techniques. For these firms, SAS No. 39 presents few problems. But for the practitioner who has relied on nonstatistical sampling where it was previously not necessary to quantify the amount of sampling risk, SAS No. 39 has effectively changed the manner in which such auditors sample. For example, the Appendix of the Statement provides a planning model that illustrates how the auditor may relate the risk of incorrect acceptance for a particular substantive test of details, to his evaluations of both the internal accounting control system and the effectiveness of any other substantive tests, such as analytical review procedures, related to the same specific audit objective.

The model is UR = IC X AR X TD. The equation can be restated to assist the auditor in planning an acceptable level of risk of incorrect acceptance (TD) after the determination of the acceptable levels of (1) ultimate risk (UR), (2) the risk of undetected error due to internal accounting control failure (IC), and (3) the risk of failing to detect errors by other substantive tests directed toward the same specific audit objective (AR).[54] The restructured form of the model's equation is TD = UR / (IC X AR). To use this

---

[54] Statistical Sampling Subcommittee, Audit and Accounting Guide: Audit Sampling, (New York: AICPA, 1983), Appendix F.

model in a nonstatistical sampling application, the auditor relies upon his professional expertise and seasoned experience to determine an acceptable level of ultimate risk (UR) and subjectively quantifies his judgment of the risks associated with IC and AR.[55]

Although this model is not intended to be used as a mathematical formula, the auditor who uses a nonstatistical sampling approach may find it useful when relating subjective evaluations of the model's factors. For example, assume that the auditor is planning a nonstatistical sampling application to test the propriety of his client's accounts receivable book value. Drawing upon his professional judgment, he may subjectively determine that a 5 percent level of ultimate risk is appropriate. Furthermore, he may subjectively evaluate the client's internal control system as warranting no reliance and thus, a 100 percent risk that the system of internal control would fail to detect material errors is proper. The auditor may also subjectively decide that his analytical review procedures and other relevant auditing procedures are only moderately effective. In this case, the auditor might assign a 50 percent risk to analytical review procedures failing to detect material errors. After he has rendered these subjective quantifications, the auditor can

---

[55] Ibid, Appendix F.

then rely upon the model to furnish him some insight as to the level of risk of incorrect acceptance that might be appropriate for his nonstatistical sampling application. Using the data in this specific example, the auditor's risk of incorrect acceptance would be determined as follows:

$$TD = UR/(IC \times AR)$$
$$TD = .05/(1.0 \times .50) = .10$$

Given the risk levels in this example, the model indicates that the auditor has an allowable risk of incorrect acceptance of 10 percent. Consequently, the auditor would then select a sample size that yields a 10 percent risk that the sample supports the conclusion that the recorded account balance is not materially misstated when in fact it is. It should be cautioned that the model is not intended to be a mathematical formula that encompases all the factors that should be considered by the auditor in his determination of sample size. For example, in his selection of the number of items to be tested for a particular test of details, the auditor should consider the allowable risk of incorrect acceptance, the tolerable error, and the characteristics of the population.[56] An auditor applies professional judgment to relate these factors in determining the appropriate sample

---

[56] Audit Standards Board, "Audit Sampling," Statement on Auditing Standards No. 39, (New York: AICPA, 1981), paragraph 23.

size.[57]

There is perhaps no better example of the influence that statistical sampling concepts have had on nonstatistical sampling applications than SAS No. 39, "Audit Sampling." To those auditors who use statistical sampling techniques, the concepts that are enumerated in SAS No. 39 are well known and present little or no problem. But for the practitioner who uses nonstatistical sampling, the promulgation of SAS No. 39 has effectively changed the manner in which such auditors perform audit sampling. For example, not only must the guidelines as established by the Statement be adhered to in accordance with the AICPA's Code of Ethics (Rule 202), but, the auditor's compliance with the Standard must be documented in his working papers. Since SAS No. 39 is equally applicable to statistical and nonstatistical sampling, all auditors must now be familiar with certain sampling concepts and be able to justify the determination of the sample size used, how the sample was selected and how the sample results were interpreted.[58]

---

[57] Ibid, para. 23.

[58] Bailey, Larry P., "Impact of SAS-39 on Nonstatistical Sampling," The CPA Journal (June 1982) pp. 38-47.

2.8 CHAPTER SUMMARY

Any discussion of the evolution of audit sampling should be within the context of changes in audit objectives. This is because a change in audit objectives generally precipitates a change in techniques. Auditing from the days of the Roman Empire to the advent of the Industrial Revolution, had as its primary objective the prevention and detection of fraud. As for audit techniques and procedures, 100 percent detailed verification of transactions against supporting documents was the well-established norm. The technique of audit sampling--or test checking, as it was once called had not yet emerged.

With the start of the Industrial Revolution in England and the resultant growth of voluminous enterprises, it became readily apparent to the auditor that an examination of every transaction in the period covered by the financial statements was inefficient and impractical. Thus, the auditor's only recourse was the implementation of a sampling application, that would provide him with reasonable assurance that the financial statements examined by him were materially correct. Unusual about this significant change in audit techniques was the fact that it was not the result of an earlier change in audit objectives. Although testing had started to be performed by the start of the twentieth century, the primary objective of auditing both in the United States and abroad, continued to be the detection and prevention of fraud. It was not until the landmark McKesson & Robbins case of 1938, that the audit profession as a whole, renounced the primacy of the fraud objective. The principal audit objective had become the auditor's determination of the fairness of the financial statements' representations.

Sampling is the process of selecting a sample from a larger group of items (called a field or population), and using the characteristics of the sample to draw inferences about the characteristics of the entire field of items. When the auditor performs sampling as one of his audit procedures, he may select either nonstatistical (judgmental) or statistical sampling or a combination of both. Prior to the 1950's, the sampling that was performed in auditing, was primarily nonstatistical in nature. A nonstatistical sample is the selection of items purely on the basis of the auditor's experience and professional judgment. Samples selected judgmentally are not supported by statistical laws of probability. As a result, the auditor is not able to objectively quantify and control the sampling risk that is associated with examining only a portion of the population items.

While nonstatistical sampling is based on the auditor's professional judgment, statistical sampling is rooted in the mathematically proven laws of probability. Accordingly, statistical sampling provides an objective manner for the auditor to (1) design an efficient sample, (2) measure the sufficiency of the evidence obtained, and (3) evaluate the sample results.[59] While the advantages identified with a statistical sampling application are quite pervasive, SAS No. 39 specifically states that ...."either a statistical or nonstatistical approach to audit sampling, when properly applied, can provide sufficient, competent evidential matter in accordance with the third standard of field work."[60] Cost-benefit considerations generally dictate the type of sampling approach that the auditor applies. Nonstatistical sampling may be less costly than statistical sampling, but the benefits from statistical sampling may be significantly greater than nonstatistical sampling.[61]

Regardless of the sampling application that the auditor selects, statistical or nonstatistical, his use of professional judgment is not minimized. For example, the auditor's sampling method does not directly influence how he evaluates

---

[59] Statistical Sampling Subcommittee, Audit and Accounting Guide: Audit Sampling Guide, (New York: AICPA, 1982), p. 11.

[60] Auditing Standards Board, "Audit Sampling," Statement on Auditing Standard No. 39, (New York: AICPA, 1981), paragraph 04.

[61] Kell & Ziegler, Modern Auditing, p. 182.

the competence of the evidence that is procured. In short, the auditor's use of professional judgment is not reduced by the selection of a particular sampling approach.

With the growth in large-scale, complex business entities, the role of sampling in professional practice has also continued to increase. Furthermore, this growth in audit sampling has spurred efforts in the profession to develop sampling guidelines that will enhance the quality of all sampling applications, particularly those that are nonstatistical. The culmination of the profession's efforts was SAS No. 39, "Audit Sampling." SAS No. 39 is quite explicit in its identification of those factors that should be considered by the auditor in his sample size determination. For the auditor who uses statistical sampling, compliance with the Statement is little problem, since he explicitly considers such factors by their inclusion in mathematical models. But for the auditor who will continue to employ a nonstatistical sampling application, SAS No. 39 will have its greater impact. In the design of nonstatistical sample sizes, the auditor must now implicitly consider the factors that SAS No. 39 enumerates. Thus, if sample sizes are to be both audit effective and audit efficient, the auditor should understand the relationship between each of the factors and sample size.

The evolution of audit sampling is characterized by many significant events, not the least of which is SAS No. 39. The Statement has effectively changed the manner in which auditors perform nonstatistical sampling. Because SAS No. 39 is couched so heavily in statistical concepts, there is concern in the profession that the Statement will unduly influence auditors to adopt a statistical sampling approach.

Chapter III

LITERATURE REVIEW

This study has as its primary objective the determination of how the typical auditor makes nonstatistical sample size decisions in substantive tests of details. That is, it seeks to identify what factors are used by the typical auditor in making nonstatistical sample size judgments in substantive tests. In pursuit of this objective, this study will utilize the methodology of Brunswik's lens model to provide mathematical representations of the typical auditor's judgment process.

This chapter reviews the research dealing with Brunswik's Lens Model as a conceptual framework for modeling human judgment under conditions of uncertainty. Judgment modeling is concerned with individuals' utilization of information when they make judgments about the current state, or predictions about some future state, of a variable of interest.[62] Those studies in accounting and auditing that are based on the methodology provided by the lens model will be reviewed and described.

---

[62] Ashton, Robert H., Human Information Processing in Accounting, Studies in Accounting Research No. 17 (Sarasota, Fla.: American Accounting Association, 1982), p. 13.

Prior to surveying the research that has utilized Brunswik's Lens Model, this chapter will review the authoritative literature in the auditing area, that presently influences the practitioner's sampling approach. In this context, the impact of Statement on Auditing Standard (SAS) No. 39, "Audit Sampling," and the related Audit Sampling Guide will be assessed.

3.1 AUTHORITATIVE LITERATURE

SAS No. 39, which supersedes sections 320A and 320B of SAS No. 1, deals with audit sampling, which is the application of an audit procedure to less than 100 percent of the items within an account balance or class of transactions for the purpose of evaluating some characteristic of the balance or class.[63]

As noted in the previous chapters, the Statement provides guidance for planning, performing, and evaluating the two general approaches to audit sampling: nonstatistical and statistical. Furthermore, SAS No. 39 applies to audit sampling both in compliance and substantive audit procedures. Because this study focuses on the auditor's nonstatistical sample size decisions in substantive tests of de-

---

[63] Auditing Standards Board, "Audit Sampling," Statement on Auditing Standards No. 39, (New York: AICPA, 1981), para. 01.

tails, it is appropriate in this chapter to review the application of the pronoucement within this context.

The purpose of substantive tests is to gather evidence as to the validity and propriety of the accounting treatment of transactions and balances. Sampling applications for substantive tests either, (1) ascertain the reasonableness of an account's recorded book value, or (2) estimate the total dollar value of some specific account.

When the auditor uses nonstatistical sampling in substantive tests, his determination of sample size is highly subjective. SAS No. 39 stipulates that the auditor should consider the following factors when he determines the sample size:

(1) Tolerable Error

(2) Allowable Risk of Incorrect Acceptance

(3) Population Characteristics[64]

Although all of the factors are highly judgmental, the auditor applies his professional judgment and seasoned experience to relate these factors in determining the appropriate sample size.

---

[64] Ibid, para. 23.

3.1.1 Tolerable Error

Tolerable error has an inverse effect on the auditor's sample size decisions. SAS No. 39 defines tolerable error as the maximum "monetary error in the related account balance or class of transactions (that) may exist without causing the financial statements to be materially misstated."[65] The Statement continues on to say that, "tolerable error is a planning concept and is related to the auditor's preliminary estimates of materiality levels in such a way that tolerable error, combined for the entire audit plan, does not exceed those estimates."[66]

Tolerable error is directly related to the concept of materiality. SAS No. 22, "Planning and Supervision," requires that the auditor consider among other things, preliminary estimates of materiality levels when he plans his audit.[67] Materiality may be thought of as the maximum monetary error that may exist in the financial statements without causing them to be materially misstated. The auditor's preliminary estimate of materiality for the financial statements taken as a whole influences the nature, timing and extent of

---

[65] Ibid, para. 18.

[66] Ibid, para. 18.

[67] Auditing Standards Board, "Planning and Supervision," Statement on Auditing Standards No. 22, (New York: AICPA, 1978).

audit procedures for particular account balances and classes of transactions.[68] The auditor in making his preliminary estimate of materiality, would ordinarily consider at a minimum the following factors: (1) the client's total assets, equity, sales and average earnings, and (2) the nature of the client's operations and related transactions.[69] Once the auditor has developed his preliminary estimate of materiality, he should then allocate this estimate to the various subsystems of the financial statements. In other words, the auditor dissects the total materiality amount into individual 'allowable errors' that pertain to specific account balances. The tolerable error that SAS No. 39 speaks of is analagous to the concept of a subsystem's 'allowable error.' Unless the auditor is able to constructively use his preliminary estimate of materiality in a manner so described, he faces the possibility of performing an audit that is ineffective and/or inefficient. As there is an inverse relationship between tolerable error and sample size, so too is there an inverse relationship between materiality levels and the scope of appropriate audit procedures.

---

[68] Zuber, Geo., Elliott, Robt., Kinney, Wm. R., Jr., and James J. Leisenring, "Using Materiality in Audit Planning," Journal of Accountancy (March 1983) p. 46.

[69] Ibid, p. 43.

Tolerable error as discussed in SAS No. 39, can be thought of as including two elements: (1) the auditor's expectation of likely error that will remain uncorrected in the financial statements after the audit is complete, and (2) an allowance for the risk of possible further error that might not be detected or indicated by the audit procedures.[70]

3.1.2 Allowable Risk of Incorrect Acceptance

The allowable risk of incorrect acceptance (TD) is a function of three factors: (1) Allowable ultimate risk, which is defined as the risk that there is a monetary error greater than tolerable error in the balance or class that the auditor fails to detect (UR), (2) the auditor's assessment of the risk that, given that errors equal to tolerable error occur, the system of internal accounting control fails to detect them, whether because of poorly designed controls or lack of compliance (IC), and (3) the auditor's assessment of the risk that analytical review procedures and other relevant substantive tests would fail to detect errors equal to tolerable error, given that such errors occur and are not detected by the system of internal accounting control (AR).[71]

---

[70] Ibid, p. 46.

The three factors can be presented in mathematical form to assist the auditor in obtaining an understanding of an appropriate risk of incorrect acceptance for a substantive test of details as follows:

$$TD = UR / (IC \times AR)$$

The allowable risk of incorrect acceptance is the risk that, given that the recorded book value is materially misstated, the auditor's other substantive tests of details will fail to detect this fact. In general, as the auditor is able to place more reliance on the system of internal accounting control (IC) and his analytical review procedures (AR), he is able to accept a greater allowable risk of incorrect acceptance. Some auditors find the above model, although it is not intended to be a mathematical formula that includes all factors that may impact the determination of individual risk components, useful when planning appropriate risk levels for audit procedures to achieve the auditor's desired ultimate risk.[72]

---

[71] Auditing Standards Board, "Audit Sampling," Statement on Auditing Standard No. 39, Appendix, p. 17.

[72] Ibid, p. 16.

In practice, auditors state in advance the level of ultimate risk that they are willing to accept after assessing the risks associated with internal accounting control (IC) and analytical review procedures (AR). That is, the level of confidence that the auditor seeks to obtain that the financial statements are not materially misstated, is established at the start of the audit engagement. With his ultimate risk level determined, as well as the risks identified with internal accounting control and analytical review, the auditor then utilizes the model to derive his allowable risk of incorrect acceptance (TD).

For example, assume that the practitioner has assessed his level of ultimate risk at 1 percent at the beginning of the audit. Furthermore, after evaluating the client's lack of internal accounting controls, he believes there is a 100 percent risk of the 'system' failing to detect a material error. Assume that the auditor also establishes a 50 percent risk of analytical review procedures failing to detect a material error. At this point, the model can be employed by the auditor to determine his allowable risk of incorrect acceptance:

$$TD = UR/(IC \times AR)$$

$$TD = .01/(1.0 \times .5) = .02$$

If the auditor is to achieve his level of ultimate risk, then his risk of incorrect acceptance (TD) cannot exceed 2 percent.

If after having performed his tests of compliance, the auditor evaluates the client's internal control system as weak, then the allowable risk of incorrect acceptance that he can accept will accordingly be reduced. The auditor's reduction of his allowable risk of incorrect acceptance will lead to a larger sample size for his substantive tests of details. In effect, the auditor adopts a substantive auditing approach, since the weak internal control system affords him little or no reliance. Conversely, the greater the auditor's reliance on analytical review procedures and other substantive tests, the greater the allowable risk of incorrect acceptance that can be accepted.

3.1.3 Population Characteristics

As the auditor sets about to determine an appropriate sample size for a substantive test, he should subjectively consider the dispersion (variation) of the amounts in the population items. As a general rule, there is a direct relationship between sample size and the population variation. That is, as the variation increases the appropriate sample size generally increases.

As an initial step in his subjective assessment of the population's variation, the auditor should review the population for those items that he considers significant, and thus should be audited 100 percent. The remaining items in the population should then be separated into relatively homogeneous groups, to minimize the effect of the variation of amounts and thereby reduce the sample size. Common bases for stratification for substantive tests are, for example, the recorded amount of the items, the nature of internal accounting controls related to processing the items, and special considerations associated with certain items (for example, portions of the population that might be more likely to contain errors).[73] The primary benefit of stratification is generally an increase in the efficiency of the sampling plan as a result of the reduction in the variation of the amounts in the groups. That is, the sample size required to separately test the individual groups, is typically not as large as the sample that would be appropriate for the population as a whole.

After having stratified the population, the auditor selects separate samples from each group and combines the results for all groups in reaching an overall conclusion about

---

[73] Statistical Sampling Subcommittee, Audit and Accounting Guide: Audit Sampling, (New York: AICPA, 1982), p. 55.

the population.[74] While the number of items in the population is usually ignored by the auditor in attribute sampling, such is not the case in substantive tests of details.

A summarization of the concepts as they pertain to audit sampling in substantive tests of details, is depicted in the flow chart of Figure 1 (p. 82). The appropriate paragraphs of SAS No. 39 are cross-referenced to the elements of the flow chart. An auditor applies his professional judgment as he seeks to relate these elements (factors) in determining the appropriate sample size.

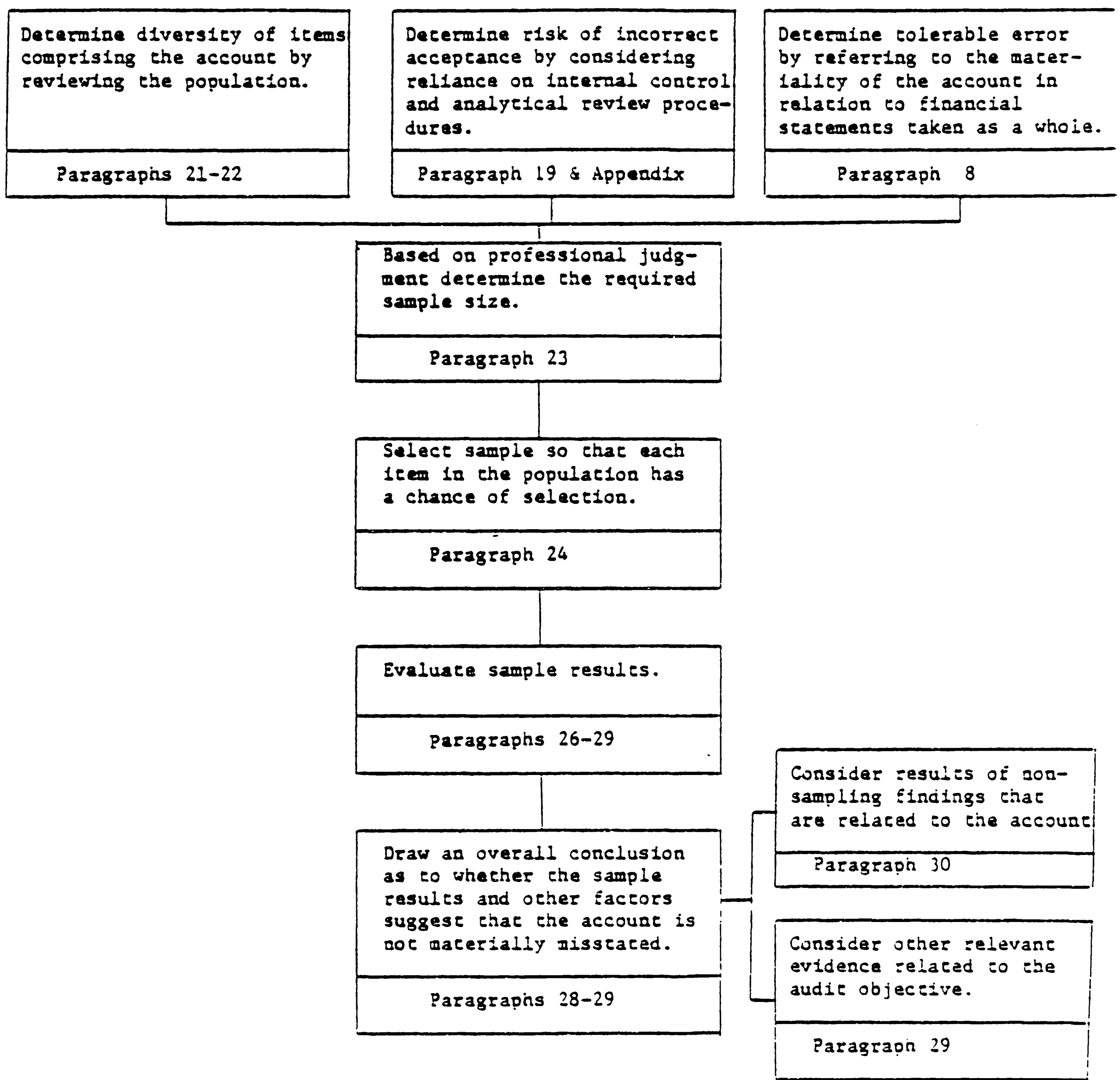
The Audit Sampling Guide prepared by the Statistical Sampling Subcommittee of the AICPA, was designed to assist the auditor in applying audit sampling in accordance with SAS NO. 39.[75] As such, the Guide presents recommendations to the auditor regarding his application of generally accepted auditing standards to audits involving audit sampling methods. In a practical sense, the Guide provides assistance to the auditor in his use of nonstatistical and statistical sampling in auditing. Since the terms and concepts of SAS No. 39, which have been explained in this study, are consistent with those enumerated in the Guide, they will not be developed further. For the auditor who is using the Guide to

[74] Ibid, p. 55.

[75] Statistical Sampling Subcommittee, Audit and Accounting Guide: Audit Sampling, (New York: AICPA, 1983), p. 7.

assist him in a nonstatistical sampling approach, Chapters 1, 2, and 3 (sections 1 and 2) provide guidance.

Flow Chart--Sampling in Substantive Tests of Details *



Bailey, Larry P., "Impact of SAS-39 on Nonstatistical Sampling," The CPA Journal (June 1982), p. 46.

FIGURE 1

Flowchart of SAS No. 39 Concepts

3.2 LENS STUDIES IN ACCOUNTING AND AUDITING

It was not until the early 1970's, that research which addressed the auditor's professional judgment emerged in the accounting literature. Since that time, studies which deal with auditor decision making under conditions of uncertainty, have been a popular form of research. Joyce and Libby cite three primary reasons for the sizeable literature that exists on the topic of auditor judgment today:

> First, the dramatic increase in litigation against auditors in the early 1970's accentuated the importance of improving the effectiveness of audits. Second, as the pressures of litigation began to subside somewhat in the late 1970's, competition among auditors for clients increased. This has led to an increased emphasis on improving the efficiency of audits. Finally, a conceptual framework for studying and evaluating auditor judgment under uncertainty-- Brunswik's Lens Model--became known to audit researchers. The increased concern about the quality of audit decision making and the existence of a mature, experimentally-based paradigm well suited to its study has been responsible for the rapid growth of the audit judgment literature.[76]

The relevant literature regarding the auditor's professional judgment is popularly known as human information processing (HIP) research in accounting.[77] The objectives of HIP research involve understanding, evaluating, and improv-

---

[76] Joyce, Edward J. and Robert Libby, "Behavioral Studies of Audit Decision Making," Journal of Accounting Literature, (Vol. 1, 1982) p. 103.

[77] Ibid, p. 103.

ing decision making as it relates to accounting.[78] HIP research in accounting falls within the boundaries of behavioral accounting research, which has been defined as ".....the study of the behavior of accountants or the behavior of non-accountants as they are influenced by accounting functions and reports.[79]

HIP research has been well serviced by two distinct paradigms: (1) the lens paradigm (which this specific study will be based on), and (2) the subjective expected utility (SEU) paradigm. Both paradigms are well-suited for modeling decision making under conditions of uncertainty. Basically, the lens paradigm focuses on the construction of linear models as mathematical representations of auditors' judgment policies. It further emphasizes the accuracy of intuitive decisions relative to those prescribed by formal models, the extent to which the intuitive decisions of the auditors agree (consensus), as well as offering several means whereby both judgment accuracy and agreement may be improved.[80] Si-

---

[78] Ashton, R. H., Human Information Processing in Accounting, Studies in Accounting Research No. 17 (Sarasota, Fla.: American Accounting Association, 1982), p. 6.

[79] Hofstedt, T. R., and J. C. Kinard, "A Strategy for Behavioral Accounting Research," The Accounting Review (January 1970), p. 43.

[80] Ashton, Robert H., Human Information Processing in Accounting, Studies in Accounting Research No. 17 (Sarasota, Fla.: American Accounting Association, 1982) p. 6.

milar to the lens paradigm, the SEU model is also concerned with modeling decision making under uncertainty. Unlike the lens paradigm, the SEU focuses on the auditor's assessment of probabilities over possible states of nature, the auditor's use of probabilities in decision making, and the revision of probabilities in accordance with Bayes' Theorem.

As noted previously, both of these paradigms are well-suited to provide the researcher with evidence as to 'how' information is utilized by the auditor in decision making. Furthermore, by comparing the auditor's actual judgments with those that would be prescribed by a normative model, both paradigms also enable the researcher to assess the accuracy of decision making. Moreover, each of these paradigms have alleviated to a significant extent, the two major limitations of earlier behavioral accounting research-- the lack of theoretical frameworks and inadequate quality control.[81]

For the most part, the research which has dealt with the modeling of auditor judgment has emphasized judgment accuracy and judgment agreement (consensus). Briefly, accuracy within the scope of the lens paradigm, can be defined as the correlation between the individual's judgments or predictions over a series of trials and those criterion values that would be prescribed by some formal model. In a similar

---

[81] Ibid, p. 7.

context, agreement can be defined as the correlation between the all possible pairs of individual's judgments over a series of trials.

Because this particular study will utilize the methodology of the lens paradigm, it is appropriate in this chapter to review the more significant experimental studies in accounting that have been based on the lens paradigm. The studies that will be described here-in address many diverse types of issues, for example, internal control evaluation, materiality judgments, bankruptcy prediction.

3.2.1 Internal Control Evaluation

The most popular topic that has been studied in the modeling of auditor judgment has been the auditor's evaluation of the strength of various fictitious internal control systems. The research within this subject matter has focused on the agreement (i.e., consensus) among the auditors' internal control assessments over a series of cases. Perhaps a limitation of the studies in this specific area, has been their inability to assess the accuracy of auditors' judgments. This is a result of the fact that there presently exists no definitive criteria for establishing the optimum internal control system.

Despite the fact that these studies have not been able to evaluate the accuracy of auditor decision making, judgmental consensus is an issue of vital importance to the profession. The importance of judgmental consensus is underscored when one considers, for example, the in-house training programs of public accounting firms and the profession's emphasis on continuing professional education. A more recent example of the profession's concern for judgmental consensus is SAS No. 39. The Statement, particularly as it applies to nonstatistical sampling, provides guidance with the aim of reducing the diversity that presently exists in practitioners' judgmental sample size decisions.

The pioneering study on the subject of internal control evaluation is Ashton (1974).[82] In Ashton (1974), thirty-two fictitious payroll internal control cases were presented to each of sixty-three practicing auditors. Ashton chose the payroll subsystem because of the risk which is so readily identified with the system, and also because the "characteristics of adequate payroll internal control should be familiar to all auditors."[83] Each case contained six questions (factors) dealing with features of internal control that had

---

[82] Ashton, R. H., "An Experimental Study of Internal Control Judgments: Replication and Extension," Journal of Accounting Research (Spring 1980) pp. 1-15.

[83] Ibid, p. 147

been pre-answered yes or no by the author. An affirmative response to a question indicated a strength in the system; whereas, a negative reply represented a deficiency. The primary purpose of Ashton's study was to ascertain the degree of consistency in practitioners' internal control evaluations. In pursuit of this objective, Ashton required each of his subjects to evaluate each system's overall strength according to a six-point scale which ranged from "extremely weak" to "strong."

Ashton was also concerned in assessing the self-insight that his subjects displayed in their judgment process. A measure of self-insight was obtained by initially having the auditors, upon completion of the experimental exercise, distribute 100 points (subjective weights) to the six factors according to the relative importance the auditor feels he has placed on each factor. The correlation between the statistically derived regression weights and the weights that have been subjectively assigned by the auditors is an indication of auditor self-insight.

The results of Ashton's study provide evidence that the judgments of individual auditors were highly consistent with those of other auditors. Judgmental consensus had an average Pearson Product Moment correlation of .70. The correlation coefficients ranged from the low 40s to the high 80s. Ashton also assessed the agreement of auditors' internal control evaluations over time (i.e., stability). For a majority of the initial sixty-three subjects, Ashton replicated the experimental exercise between forty-three and ninety-four days later to be able to assess the stability of auditors' decisions. Again, using correlational analysis, the findings were quite impressive. On the average, the correlation between each auditor's judgments on the two administrations of the exercise was .81. Of the six dichotomously scaled factors, two that pertained to separation of duties were the most influential in the auditors' overall internal control evaluation. Finally, the auditors exhibited remarkable self-insight into their judgment process with a mean correlation of .89.

Ashton's (1974) original study was replicated to a great extent by Hamilton and Wright (1977).[84] With seventeen practitioners serving as subjects, Hamilton and Wright presented to each the thirty-two fictitious payroll internal control cases that were used in Ashton's earlier study. As a slight modification to Ashton's experiment, two of the six original factors were deleted, while the two factors that pertained to separation of duties were reconstructed and

---

[84] Hamilton, R. E. and W. F. Wright, "The Evaluation of Internal Controls Over Payroll," unpublished manuscript, Stanford University, 1977.

presented to the participants as three factors.

Again, the findings were particularly strong. The average correlation between the judgments of all possible pairs of auditors (judgmental consensus) was .66. The three factors that concerned the separation of duties, explained on average, 51 percent of the variance in auditors' internal control evaluations. Again, as was the finding in the Ashton (1974) study, the researchers found that auditors possess particularly strong insight into their judgment process. The mean correlation of 'self-insight' was .87.

Ashton and Brown (1980), with some modifications, also replicated Ashton's pioneering work.[85] To the original six internal control factors, two additional questions one of which concerned the separation of duties were added, to produce a more extensive decision task. Moreover, each of the thirty-one auditors in the study was required to evaluate internal control systems for one-hundred and twenty-eight hypothetical cases. In addition to the lot of one-hundred and twenty-eight, each subject also received thirty-two repeat cases. The judgments of the auditors over the thirty-two repeat cases would permit the authors to assess judgment stability. Ashton and Brown also evaluated the auditors in-

---

[85] Ashton, R. H. and P. R. Brown, "Descriptive Modeling of Auditors' Internal Control Judgments: Replication and Extension," Journal of Accounting Research (Spring 1980), pp. 269-77.

sight into their judgment process.

The findings of Ashton and Brown approximated the results of the studies cited earlier in this chapter. The mean correlation of judgment consensus was a moderate .67, which was slightly higher than what the Hamilton & Wright study found (.66), yet somewhat lower than Ashton's original study (.70). The three factors that related to separation of duties features accounted for approximately 51 percent of the variance in the auditors' judgments. Similar to the results in the previous studies, auditors exhibited particularly keen insight into their judgment process. The average correlation of self-insight was .86

Ashton and Kramer (1980) replicated Ashton (1974) with accounting students serving as surrogates for practicing auditors.[86] Each of thirty auditing students completed Ashton's original experiment involving thirty-two cases and six factors. The study's findings can perhaps be best explained by the students' lack of professional expertise. The average Pearson correlation representing the students' judgmental consensus was .66, which was slightly lower than the auditors' results of .70. Furthermore, the two factors which related to the separation of duties only explained 37 per-

---

[86] Ashton, R. H., and S. S. Kramer, "Students as Surrogates in Behavioral Accounting Research: Some Evidence," Journal of Accounting Research (Spring 1980), pp. 1-15.

cent of the variance in students' judgments. In contrast, auditors apparently attach more significance to the separation of duties, as the factors that pertained to these internal control features, accounted for 51 percent of the variance in auditors' judgments. The students' insight measure, although less than that of auditors, was a moderate .77

In a 1979 study, Reckers and Taylor were interested in assessing what impact the 'length of auditing experience' would have on the judgmental consensus of both practicing auditors and auditing professors.[87] In a research design that was rather complex, thirty practicing auditors and forty professors of auditing, were each asked to evaluate the reliance of internal control systems on five fictitious cases. Each of the cases contained thirty-six factors representative of internal control features. The subjects were asked to rate the reliability of each of the five systems on a range from 0 to 100 percent.

The mean consensus for auditors and professors was .16 and .13 respectively. A possible explanation for these low correlations may be the difficulty that was possibly experienced by the subjects in assimilating the effect of thirty-

---

[87] Reckers, P. M. J., and M. E. Taylor, "Consistency in Auditors' Evaluations of Internal Accounting Controls," Journal of Accounting, Auditing and Finance (Fall 1979), pp. 42-53.

six factors on the reliability of internal control. Furthermore, these results infer that the high consensus of the auditors in the studies cited earlier, was due to the fact that the cases contained only six factors. However, the authors did obtain evidence that the level of auditing experience can impact the consensus of payroll internal control evaluations. For the group of practitioners who exceeded the median experience level of 7.5 years, the mean value of consensus was .36. Conversely, those practicing auditors whose level of professional experience was less than the median amount, their average consensus was .14. The results are interesting in the sense that they apparently underscore the importance of practical experience in judgmental consensus. For this specific experimental exercise, the mean consensus (.14) of practicing auditors with less than the auditor median experience level, was greater than the average correlation (.13) between all pairs of auditing professors over all of the cases. These results suggest that, professional experience may have important implications on the judgmental process of individuals.

Joyce (1976) had thirty-five practicing auditors evaluate several fictitious accounts receivable internal control cases, each of which contained five dichotomously scaled internal control factors.[88] For each case that the auditors reviewed, they were to indicate the number of man-hours they would allocate to five distinct categories of audit procedures (e.g., review of year-end sales cutoff, confirmation procedures). Joyce sought to evaluate the consistency of auditors' judgments on a basis of time estimates in audit program planning. With specific reference to Ashton's pioneering study (1974), Joyce theorized, that although auditors exhibited high judgmental consensus in their overall evaluations of internal control systems, they may demonstate less consistency if pressed to allocate audit hours to five substantive testing procedures.

As were the findings in the previous studies dealing with the evaluation of internal control, the one factor that pertained to the separation of duties was important to the auditors' judgments. On the average, the 'separation of duties' factor accounted for approximately 28 percent of the judgment variance. Contrary to the earlier results (Ashton (1974), Hamilton and Wright (1977), Ashton and Brown (1980)), the consensus among the auditors on their assessment of the number of audit hours to be allocated to specific audit procedures was a low .37.

---

[88] Joyce, E. J., "Expert Judgment in Audit Program Planning," Studies in Human Information Processing in Accounting, Supplement to Journal of Accounting Research (1976), pp. 29-60.

In yet another variation in research design within the context of internal control evaluation, Mock and Turner (1979) studied the effect of internal control changes on the auditors' judgments.[89] Seventy-three practicing auditors were given only one case which dealt with a company's revenue cycle and were asked to make judgments relative to the appropriate sample sizes for four substantive audit tests. For half the auditors, their case reflected a marked improvement in internal controls; while the other half received a case in which internal controls only slightly improved.

The results of Mock and Turner's study show that auditors whose case reflected a substantially improved internal control system recommended significantly smaller substantive test sample sizes. In fact, these results were consistent over the four substantive audit tests. With respect to judgmental consensus, their findings are consistent with those of Joyce (1976). A considerable amount of inconsistency was found in the sample size recommendations among the subjects. This specific study also provided evidence of the effect of an anchor on the subjects' sample size recommendations. That is, each of the seventy-three auditors was given the present

[89] Mock, T. J., and J. L. Turner, "The Effect of Changes in Internal Controls on Audit Programs," Behavioral Experiments in Accounting (Ohio State University, 1979), pp. 277-302.

year's planned sample size in his case narrative. In a subsequent administration of the same cases that also involved practicing auditors, no anchor was given. For the group involved in the initial study, the results suggest that the auditors were influenced by the anchor in the recommendations of this year's sample size.

To briefly summarize, the evaluation of internal control has been the most popular topic studied in the accounting literature that deals with the modeling of auditor judgment. For the most part, the research results show a high level of consistency (i.e., consensus and stability) in the judgments among practicing auditors. This finding is particularly true in those experimental tasks that are relatively straightforward and simple. As task complexity increases, judgmental consensus has been shown to decline. Furthermore, those factors or features of internal control that are associated with the separation of duties are typically able to explain most of the variance in the auditors' judgments. For those studies that evaluated the subjects' insight into their judgment process, the fingings were conclusive. Auditors display remarkable insight into identifying those factors that are most influential to their judgment process. Finally, there is some evidence to indicate that, length of professional experience can contribute to judgmental consensus. This is especially true in those experiments where the decision task is realistic and/or complex.

The review of studies in the accounting literature that are based on the lens paradigm, will focus now on those studies that have used the model in the prediction of bankruptcy.

3.2.2 Prediction of Bankruptcy

Unlike the lens studies that dealt with the auditors' evaluation of internal control, lens studies on the topic of bankruptcy prediction permit the assessment of judgmental accuracy. Because the experimental tasks which deal with the prediction of bankruptcy are developed from actual firms' data, information as to whether a company has failed or not is readily available, and is compared to the subjects' prognostications of failure as an index of judgmental accuracy.

Libby (1975) used forty-three bank officials to predict whether each of sixty firms had failed or not within a three-year period.[90] To assist the decision makers in their judgment task, five financial ratios (e.g., net income/total assets, current assets/sales, current assets/current liabilities) that were based on each firm's data were made availa-

---

[90] Libby, R., "Accounting Ratios and the Prediction of Failure: Some Behavioral Evidence," Journal of Accounting Research (Spring 1975), pp. 150-161.

ble. The bank officers were told that of the sixty firms, half had failed within the three-year period. In order to assess judgment stability, ten repeat firms (half of which failed) were also given to the subjects. The ten repeat firms "were placed as every fourth firm following the first thirty cases to conceal the fact that they were repeat firms."[91] A measure of performance for the bank officials was developed by granting one point for each correct prognostication, while subtracting a point for each incorrect prediction. Of the forty-three respondents, only three had an accuracy index within the random range. The remaining forty on average, successfully predicted 44.4 of 60 firms. The binomial test[92] was used to assess the chances of obtaining at least the actual proportion of correct prognostications given random assignment. In this specific study, there was less than a 5 percent chance of obtaining at least thirty-seven correct responses as a result of random prediction. Perhaps because of the predictive strength of the financial ratios, length of professional experience did not correlate with judgmental consensus. Thus, it is apparent that the five ratios (factors) served a very useful purpose.

---

[91] Ibid, p. 482.

[92] Siegel, S., Nonparametric Statistics for the Behavioral Sciences (New York: McGraw-Hill, 1956).

Consistent with other lens studies in accounting, judgmental consensus and stability were also assessed in the Libby study. The mean consensus among the bank officials was a moderate .48. That is, after considering all pairs of subjects' predictions over sixty firms, the bank officers agreed on average, forty-eight times. With respect to judgmental stability, an overwhelming majority of the forty-three respondents rendered the same prediction at least 80 percent of the time.

In a partial replication of the Libby Study (1975), Casey (1980) had forty-six bank officials predict whether each of thirty firms had failed or not failed.[93] On average, each participant in the study held the title of assistant vice-president and had 7 years of professional experience as a loan officer. In a departure from the Libby design, the subjects were not informed of the number of firms that had gone bankrupt. Thus, the effects of subjects' prior probabilities of failure cannot be discounted in the study's results. Although one additional (i.e., total liabilities/ owners' equity) was used in conjunction with the five original, the mean accuracy of the subjects' predictions for the thirty firms was a disappointing seventeen. At the 99

[93] Casey, C. J., "The Usefulness of Accounting Ratios for Subjects' Predictions of Corportate Failure: Replication and Extensions," Journal of Accounting Research (Autumn 1980), pp. 603-13.

percent level of confidence, at least twenty-one correct responses were necessary to attain greater than chance (random) accuracy. The subjects' failure to achieve non-random accuracy, can be traced to the fact that on average, they correctly predicted the failure of only four of fifteen bankrupt companies. Conversely, for the fifteen firms that did not fail, the subjects' accuracy was particulary impressive. On average, they were able to correctly predict the nonfailure of thirteen of the fifteen non-failed firms. It can be inferred from these results that subjects not having been told how many of the firms failed, tended to be optimistic as they made their predictions. For example, approximately 76 percent (13 of 17 firms) of the bank officers' predictions represented firms that did not fail.

Casey also assessed judgmental consensus in his study and found agreement among the bank officers' judgments to be moderately strong. The average number of times on which pairwise predictions agreed was twenty-four. Subject agreements ranged from eighteen to twenty-nine firms. Intersubject agreement (consensus) was also assessed by determining the percentage of each firm's predictions that agreed. Subject agreement by firm ranged from 54 percent for two of the firms to 100 percent for nine of the firms and averaged 89 percent.[94]

In a study published in 1980, Kida had twenty-seven audit partners predict how many of forty companies would undergo going concern problems.[95] Audit partners were selected as subjects since "they were the individuals ultimately responsible for their firms' going concern qualification decisions."[96] Kida did not inform the partners that 50 percent of the forty firms did in fact experience going concern problems. A firm was considered as having going-concern problems, if any one of a limited number of events (e.g., enter receivership, enter reorganization proceedings, inability to meet interest payments) occurred within one year of the financial statement date. Similar to Libby (1975), Kida also furnished his subjects with five financial ratios to assist them in their judgment decision.

In addition to the typical indices of judgment (e.g., accuracy, consensus) that are usually assessed in lens model studies, Kida also wished to determine whether a positive association exists between auditors' qualifying attitudes and their decisions as to whether a firm will experience go-

---

[94] Ibid, p. 611.

[95] Kida, Thomas, "An Investigation Into Auditors' Continuity and Related Qualification Judgments," Journal of Accounting Research (Autumn 1980), pp. 506-23.

[96] Ibid, p. 508.

ing concern problems. The auditors' qualifying attitudes were measured by the Fishbein and Ajzen model,[97] and reflect the auditors' attitudes toward issuing unqualified-qualified- disclaimed audit opinions.

On the accuracy dimension, the audit partners performed very well. With an accuracy rate of twenty-eight or more having a less than 1 percent chance of occurring from random predictions, the twenty-seven partners attained on average, 33.2 correct responses. The range of correct responses was twenty-seven to thirty-seven, with one of the partners correctly predicting 95 percent (38 of 40) of the time. Kida also found the agreement (i.e., judgmental consensus) among auditors' judgments to be substantial. Using Kendall's coefficient of concordance (W), judgmental consensus was assessed at .765, which was significant at the .001 level. When the 'W' statistic was transformed to the average Spearman's rho, judgmental consensus was approximately .76. Finally, this study provided evidence that a positive association exists between auditors' going concern decisions and their qualifying attitudes. At the .016 level, there existed a positive correlation of .42.

---

[97] Fishbein, M., and I. Ajzen, Belief, Attitude, Intention and Behavior (Reading, MA.: Addison-Wesley, 1975).

Moriarity (1979) had two-hundred and twenty-seven students from a large mid-western university, predict whether twenty-two firms from the discount department store industry had sought protection under Chapter 10 or Chapter 11 of the Federal Bankruptcy Laws.[98] Similar to the Kida study (1979), the students were not appraised that seven of the twenty-two firms had failed. The participants in the study were presented information in one of four different ways to assist them in making their prognostications. One group was presented with selected financial statement balances (e.g., net sales, net profit, total current assets), that the participants could use to calculate financial ratios. A second group was given thirteen financial ratios (e.g., inventory/working capital, sales/working capital), that were prepared from Dun and Bradstreet's Key Business Ratios. With the remaining two groups, Moriarity used schematic faces as developed by Chernoff,[99] to communicate each firm's financial information. One of the groups to receive the Chernoff facial features (e.g., Ear Diameter, Nose Width, Mouth Curvature, Face Height) was also informed of the financial ratio that

---

[98] Moriarity, Shane, "Communicating Financial Information Through Multidimensional Graphics," Journal of Accounting Research (Spring 1979), pp. 205-24.

[99] Chernoff, H., "The Use of Faces to Represent Points in N-Dimensional Space Graphically," Technical Report No. 71 (Department of Statistics, Stanford University, December 1971).

each facial feature represented. The remaining group was only informed that the schematic faces represented a firm's financial information. The subject could discern a structural change in a firm's financial status, by noting the changes in the facial features. For example, an increase in the size of the eyes was indicative of an increase in the firm's long-term debt.

The purpose of Moriarity's study "was to determine whether unsophisticated users could use faces to make better financial decisions.[100] Moriarity theorized that those individuals who are uncomfortable with manipulating numeric information, may be able to turn to Chernoff faces to assist them in assessing a firm's financial position. Furthermore, he felt that Chernoff faces could be used to evaluate a firm's financial status more efficiently, particularly for those persons who are unable or unwilling to invest their time in financial statement analysis.

The study's results showed that the group of students who received the Chernoff faces with an accompanying explanation attained the highest predictive accuracy. In fact, both groups to receive Chernoff faces outperformed the other two groups who were given quantitative data. The seventy-nine students who received the financial information in the

---

[100] Moriarity, S., "Communicating Financial Information Through Multidimensional Graphics," p. 207.

form of ratios, had the dubious distinction of being the least accurate group. These findings appear to cast doubt as to the predictive ability of financial ratios, particularly when compared with other modes of data presentation.

To recapitulate, those studies that have dealt with the prediction of firm failure, have generally found subjects to be fairly accurate in their bankruptcy predictions. Of the studies reviewed, only in one case (Casey 1980) was the predictive ability of the subjects found to be wanting. For the most part, judgmental accuracy was evaluated using bank loan officers or CPA partners as subjects and key financial ratios as predictor variables. In the studies that evaluated intersubject consensus, the findings were conclusive. The average agreement between the predictions of all pairs of subjects was strong.

The lens paradigm has also been extensively used to model subjects' materiality judgments. A review of some of the more important studies that have addressed this issue will now be presented.

3.2.3 Materiality Judgments

Boatsman and Robertson (1974) used the methodology provided by the lens model to describe the process of materiality judgment formulation.[101] Thirty fictitious cases, each of which contained the same eight factors (e.g., nature of item, relationship of the item to current year income) were presented to thirty-three subjects (eighteen CPA partners, fifteen security analysts). The eight factors chosen for inclusion in the case narratives were selected on the basis of their assumed relevance to the materiality judgment process. The experimental exercise required that each subject sort the cases into one of three disclosure classifications (1) no separate disclosure, (2) footnote disclosure, and (3) line-item disclosure. The linkage of materiality to disclosure was made under the premise that, "when alternative methods of disclosure are available for a given item, the materiality of the item determines the disclosure."[102]

With thirty-three participants each evaluating thirty cases, a total of nine-hundred and ninety observations were gathered from which to model the subjects' composite judgmental process. Using multivariate discriminant analysis, the model was able to predict 625 (63 percent) of the 990 materiality decisions. All of the independent factors were statistically significant, yet only factor two (i.e., the

---

[101] Boatsman, J. R., and J. C. Robertson, "Policy-Capturing on Selected Materiality Judgments," The Accounting Review (April 1974), pp. 342-52.

[102] Ibid, p. 346.

relationship of the item to current year income) was found to have substantial explanatory power. Of the total predictive power (63 percent), factor two accounted for 73 percent. The researchers also tested the hypothesis that no significant differences existed between the judgmental processes of the CPAs and the security analysts. Grouping the disclosure decisions of the CPAs and the analysts separately, models representative of each group's judgmental process were formulated. Rankings were then assigned to each of the factors on a basis of their relative importance to each model. The Spearman rho was used to compute the correlation between the two sets of ranks. At the .05 level of significance, the correlation of .67 suggests that, both groups of professionals utilize the same independent factors to approximately the same degree.

Ward (1976) evaluated the degree of consensus among practicing auditors on the aspect of how they perceive the relative importance of factors associated with the issue of materiality.[103] Ward hypothesized that "auditors as a class are homogeneous in regard to their cognitive beliefs about the materiality construct; (however), individual differences may influence the operational perceptions or affective

---

[103] Ward, Bart H., "An Investigation of the Materiality Construct in Auditing," Journal of Accounting Research (Spring 1976), pp. 138-52.

aspects of attitudes about materiality."[104] A total eight partners and sixteen managers from three international public accounting firms acted as subjects in the study. The experimental task required each of the respondents to rank five factors (legal, technical, professional, personal, and environmental) according to their relative importance in the subjects' materiality judgments.

The results provide evidence that practitioners tend to be in significant agreement regarding the relative importance of the five factors to their materiality judgments. Judgmental consensus, as assessed by the Kendall Coefficient of Concordance was .386, ($p < .01$). The study's results also show significant consensus among the practicing auditors at the firm level.

Hofstedt and Hughes (1977) had nineteen graduate business students determine the probability that he or she would disclose the writedown of a totally owned subsidiary.[105] The disclosure alternative necessitated reflecting the loss seperately as an extraordinary item, while the nondisclosure alternative meant aggregating the loss under "Other Expenses." The researchers assumed that the probability of disclo-

---

[104] Ibid, p. 141.

[105] Hofstedt, T. R., and G. D. Hughes, "An Experimental Study of the Judgment Element in Disclosure Decisions," The Accounting Review (April 1977), pp. 379-95.

sure as evaluated by the subject, was a function of how the subject assessed materiality. The students were presented with three criteria (factors) to assist them in resolving the materiality/disclosure problem. All of the factors concerned the size of the writedown relative to the magnitude of (1) operating income, (2) all parent investments in unconsolidated subsidiaries, and (3) the net book value of the subsidiary being written down.[106]

The results showed the three factors (materiality criteria) to be statistically significant at the .05 level of confidence. A simple regression model was able to explain about 39 percent of the judgment variance. The regression model for individual subjects based on main effects only, was able to account for an average .70 of the judgment variance. The coefficient of determination for individual subjects ranged from .40 to .94. In general, the individual analyses revealed substantial differences among the judgment policies of the subjects. Of the three main factors, the 'loss as a percentage of net income' was the most important to the participants. Finally, as a measure of self-insight into their judgmental process, the subjects were asked to allocate one-hundred points among the three factors in accordance with their perceived importance. In comparison to

---

[106] Ibid, p. 383.

the objective factor weights as derived from the regression technique, the subjects tended to "overstate the importance of minor cues (factors) and to understate the importance of major cues."[107]

Moriarity and Baron (1976) sought to determine how fifteen audit partners from eight large public accounting firms made materiality decisions.[108] Unlike the previous studies that focused on materiality, this studied viewed materiality and its contributing factors (e.g., net income, asset size, earnings trend) as an ordinal rather than a nominal concept. This treatment enabled the researchers "to obtain comparative judgments from the individual auditor, thereby removing (materiality) thresholds as a source of difference in individual judgments."[109] Whereas Boatsman and Robertson (1974) employed discriminant analysis to study materiality, Moriarity and Baron used conjoint measurement techniques to derive both the functional form of the subject's decision model, and the contributory effects of three factors (net income (three-levels), firm size (three-levels), and earnings trend (two-levels)) on materiality judgments. The three factors

---

[107] Ibid, p. 387.

[108] Moriarity, S., and F. Hutton Baron, "Modeling the Materiality Judgments of Audit Partners," Journal of Accounting Research (Autumn 1976), pp. 320-41.

[109] Ibid, p. 322.

were varied in a completely crossed design to produce eighteen sets of financial statements. The item in each of the financial statements that was the basis for the subjects' materiality judgments was a before-tax decrease in earnings of $500,000. The experimental task required the auditors to rank the 18 cases according to their perception of the materiality of the earnings decrease to each case.

The analysis of the data involved the identification of the functional form (e.g., additive, multiplicative, distributive) of each subject's judgment model. For the eleven participants whose judgment models approximated an additive form, Kruskal's MONANOVA[110] was used to evaluate the importance of each factor level to the subject's overall materiality judgments. For each of the audit partners, the 'net income' factor was clearly the most explanatory factor. Thus, the higher a case's net income, the lesser was the perceived materiality of the $.5 million decrease in earnings.

To summmarize, the studies of materiality judgments have linked the construct of materiality with disclosure decisions. Such linkage has been based on the assumption that, with various forms of disclosure available for a given item,

---

[110] Kruskal, J. B., "Analysis of Factorial Experiments by Estimating Monotone Transformations of the Data," Journal of the Royal Statistical Society Series B, No. 2 (1965), pp. 251-63.

the selection of a particular method is predicated on the item's perceived materiality. The results of the research are fairly conclusive on one particular aspect; namely, the importance of net income to decisions on materiality. Aside from this finding, there appears to be little consensus among subjects on materiality judgments.

3.2.4 Other Selected Studies

In yet another application of the lens paradigm, Wright (1977) had thirty-nine second-year MBA students estimate security price changes over a one-year period.[111] Among the author's objectives was the assessment of judgmental accuracy and interjudge agreement. The experimental task required the subjects to estimate ex post price changes for sixty common stocks from one generic industry (industrial and construction machinery and metal products). To assist the subjects in their decision task, five factors (e.g., cash dividend, earnings-per-share, ratio of long-term debt to total stockholders' equity) were constructed based on each firm's actual accounting data and capital market information.

---

[111] Wright, William F., "Financial Information Processing Models: An Empirical Study," The Accounting Review (July 1977), pp. 676-89.

An analysis of the students' predictions indicates that judgmental accuracy was quite low. At least 14 correct predictions (alpha = .05) were necessary in order to achieve accuracy on a better than random basis. Fewer than 40 percent of the students exceeded the level of chance accuracy. The median accuracy correlation was .16, with a range of -.08 to .44. On the average, the five predictor factors were able to explain 34 percent of the judgment variance. The median proportion of judgment variance accounted for was .49, with a range of .13 to .79. Finally, the results indicate little intersubject agreement among the estimated price change predictions. The median judgmental consensus correlation was .27. These results are consistent with the substantial heterogeneity that was observed in individual judgment distributions.[112] For example, the students estimated mean price changes ranged from -4.63 to 9.99.

Schultz and Gustavson (1978) modeled the judgmental process of five actuaries to empirically determine which of five factors (e.g., client's financial condition, client size, accounting firm's size) they weigh most heavily in assessing the riskiness of an audit engagement.[113] Each of

---

[112] Ibid, p. 682.

[113] Schultz, Joseph J., Jr., and Sandra G. Gustavson, "Actuaries' Perceptions of Variables Affecting the Independent Auditor's Legal Liability," The Accounting Review (July 1978), pp. 626-41.

the five participants was professionally employed at the time the study was conducted with a different insurance broker that sold accountants' professional liability insurance. The experimental exercise required each of the subjects to assess "the likelihood of a valid and substantial claim being sustained by the hypothetical (accounting) firm within one year."[114] With five predictor variables (2-levels each) being manipulated in a completely crossed factorial design, a total of thirty-six cases (which included four repeats) were evaluated by each actuary. The four repetitive cases were included to enable the authors to evaluate the reliability of each subject's decisions. The primary objectives of this study were three-fold: (1) to assess judgmental consensus, (2) to determine the relative importance of the factors to the actuaries' judgments, and (3) to measure actuary self-insight. Since the cases were purely hypothetical and there were no actual malpractice claims, it was not possible in this study to evaluate the accuracy of the judgments made by the actuaries.

On the judgmental consensus dimension, the average Pearson Product-Moment correlation between the judgments for all pairs of actuaries was .12. The agreement between the judgments of each subject ranged from -.23 to .52. These

---

[114] Ibid, p. 630.

results indicate that little judgmental consensus existed. The results further indicate that each of the five actuaries relied heavily upon the five factors in rendering their judgments. For each of the subjects, "the main effects of the cues (factors) are significant at least at the .05 level, except for one instance which was significant at the .10 level."[115] When the main effects are taken together with two and three-cue interactions, a median of 94.3 percent of the judgment variance is explained. Based on the four repetitive cases that were administered to each participant, the test-retest reliability was a perfect 1.0 for each actuary. Finally, the actuaries exhibited a moderate degree of self-insight into their judgmental process. Three of the five respondents had product-moment scores of .92, .99, and 1.0. On the average, judgmental insight was .78.

Danos and Imhoff (1980) conducted a study, the primary purpose of which was to assess the impact of various factors on auditors' reasonableness judgments of financial forecasts.[116] A total of forty praticing auditors, the majority of which were at the level of partner, participated in the experimental exercise. Each of the auditors was asked to re-

---

[115] Ibid, p. 637.

[116] Danos, P., and E. A. Imhoff, "Auditor Review of Financial Forecasts: An Analysis of Factors Affecting Reasonableness Judgments," The Accounting Review (January 1982), pp. 39-69.

gister a reasonableness judgment according to a ten-point scale for each of thirty-two forecasted corporate income statements. Within each of the experimental corporate cases, five factors (two-levels) were manipulated in a one-half fractional factorial design. The five factors (e.g., accuracy of sales forecasts, accuracy of net income forecasts) were developed based on pre-test interviews that were conducted with the study's participants. Unlike the research that focused on bankruptcy predictions, Danos and Imhoff did not evaluate judgmental accuracy, "since the objective of this study was not to determine whether auditors could differentiate ex ante between accurate and inaccurate information."[117]

For each of the forty auditors two judgmental models were produced, since one-half of the experimental cases involved factual data from a brewing company, while the remaining cases were developed from the records of a men's apparel company. Based on a total of eighty judgment models being computed (forty subjects X two models each), the most significant factor was that which pertained to the "Accuracy of Net Income Forecasts." In 60 percent of the subjects' judgment models, this predictor variable was significant at .05. On average, the linear judgment models were able to ex-

---

[117] Ibid, p. 40.

plain approximately 80 percent of the judgment variance. The level of judgmental consensus was moderate. For 36 percent of the subjects' models, the "Accuracy of Net Income Forecasts" was considered to be the most significant variable (alpha < .05). Similarly, the factor which pertained to the "Accuracy of Sales Forecasts," was most significant in 31 percent of the decision models. The degree of auditor self-insight into his judgmental process was also quite strong. Twelve (30 percent) of the forty participants were able to identify the most significant factor in both of their experimentally determined decision models. Furthermore, an additional nineteen auditors (47.5 percent) correctly determined the most explanatory variable in one of their two judgment models.

Kinney and Uecker (1982), motivated to assist the AICPA's Sampling Standards Task Force in drafting a proposed SAS on audit sampling, investigated differences in auditors' evaluations of compliance sample outcomes.[118] A total of one-hundred and sixty-nine audit seniors, representing five multi-national public accounting firms, judgmentally evaluated sample results by either the fractile assessment or risk assessment techniques. The authors tested for the 'an-

---

[118] Kinney, William R., Jr., and Wilfred C. Uecker, "Mitigating the Consequences of Anchoring in Auditor Judgments," The Accounting Review (January 1982), pp. 55-69.

choring' effect of sample results in the auditors' judgmental evaluation of compliance test outcomes.

For the group who employed the fractile assessment method, the authors hypothesized that the auditors would too frequently understate the achieved upper precision limit for the population error rate. As a result, these auditors are likely to place too much reliance on internal controls thus increasing the probability of committing a type II error. Conversely, for the group who utilized the risk assessment technique, the authors hypothesized that the auditors would too frequently overstate the probability that the population rate exceeded the maximum allowable rate. Consequently, these auditors are apt to place little or no reliance on internal controls thus increasing the probability of committing a type I error. Because type II errors impact audit effectiveness, they are potentially of a more serious nature than type I errors which relate primarily to audit efficiency. Past research studies have provided evidence that individuals tend to make estimates based on some initial value (an anchor) that may be suggested by the task or may be the result of a partial computation, and then proceed to make insufficient adjustments to this anchor.[119]

---

[119] Ashton, R. H., Human Information Processing in Accounting, Studies in Accounting Reserarch No. 17 (Sarasota, Fla.: American Accounting Association 1982), p. 107.

The experimental task required each auditor to evaluate a hypothetical case's sample result, and either indicate the probability that the true population error rate is greater than 8 percent (risk assessment method), or estimate the 95th fractile on population error rate (fractile assesment). A total of eighty-six auditors evaluated their sample's results according to the fractile assessment method, while eighty-three auditors utilized the risk assessment technique. The results of the study confirmed the hypothesized effects of anchoring for the two assessment techniques. For the fractile assessment group, the null hypothesis that the proportion of type I and type II errors is the same was strongly rejected (p-value < .001). Within this particular group, a comparative analysis of the auditors' median responses with the corresponding statistical evaluation for the cases, revealed auditors' judgments to be consistently less than the statistical evaluation. The typical consequence of such judgmental evaluations is a type II error. Conversely, a greater proportion (non-significant) of type I errors was experienced by the auditors who were instructed to use the risk assessment method. Overall for the two assessment techniques, the proportion of type I and type II errors was significantly different (chi-square statistic = 13.18, p-value < .001).

The results of this research have had a direct and profound impact on the formulation of the authoritative pronoucement, SAS No. 39. As a result of the empirical evidence gathered in this study, "the Sampling Standards Task Force decided to revise its proposed SAS to exclude suggestion of the fractile assessment method."[120] Consequently, SAS No. 39 advocates the use of the risk assesment method in auditors' judgmental evaluation of sample results.[121]

3.3 CHAPTER SUMMARY

This chapter reviewed the authoritative literature on audit sampling and selectively surveyed some of the more significant accounting research studies that have been based on Egon Brunswik's lens model.

In an authoritative context, SAS No. 39 has effectively changed the manner in which practitioners perform audit sampling. This is especially true for those auditors who apply nonstatistical sampling techniques. With the promulgation of SAS No. 39, the auditor must implicitly consider the following factors when he determines the sample size: (1) tol-

---

[120] Kinney, William R., Jr., and Wilfred C. Uecker, "Mitigating the Consequences of Anchoring in Auditor Adjustments," p. 68.

[121] Auditing Standards Board, "Audit Sampling," Statement on Auditing Standards No. 39, (New York: AICPA, 1981), para., 40.

erable error, (2) allowable risk of incorrect acceptance, and (3) population characteristics. An auditor applies his professional judgment as he seeks to relate these factors in determining the appropriate sample size.

The selected lens studies in accounting that were reviewed in this chapter have addressed some subset of the following topics: (1) the relative importance of various factors to the judgment processes of individuals, (2) the accuracy of human predictions, (3) the level of judgmental consensus across individuals, (4) the stability of individuals' judgments over time, and (5) the insight exhibited by individuals into their judgment process. The studies that were surveyed came primarily from three major research areas: (1) internal control evaluation, (2) bankruptcy prediction, and (3) materiality judgments.

In the area of internal control evaluation, the research results show a high level of agreement in the judgments among practicing auditors. Moreover, those factors or features of internal control that pertain to the separation of duties, are usually able to explain most of the variance in auditors' judgments. As a group, auditors also display keen insight into identifying those factors that are most influential to their judgment process.

Those studies that dealt with the prediction of firm failure have found the consensus among individuals' predictions to be especially strong. However, the research results on the accuracy dimension appear to be a function of whether subjects, prior to commencing the experimental task, are appraised of the number of firms in the sample that actually failed.

Studies of materiality judgments have linked the construct of materiality with disclosure decisions. Fairly consistent evidence has been gathered in this area on the importance of net income to subjects' materiality judgments. Aside from this finding, there appears to be little consensus among subjects on materiality judgments.

This chapter's survey of selected behavioral studies in accounting also included research that has focused on other decision areas (e.g., security price predictions, auditors' reasonableness judgments). One study in particular (Kinney and Uecker, 1982), provided rather strong evidence that auditors who use the fractile assessment method to judgmentally evaluate sample results, are likely to place too much reliance on internal controls thus increasing the probability of committing a type II error. These findings prompted the AICPA's Sampling Standards Task Force to amend its proposed statement on audit sampling to exclude suggestion of the fractile assessment technique.

Chapter IV

THE EMPIRICAL STUDY: RESEARCH METHODOLOGY

This chapter will review the research methodology of an empirical study designed to provide evidence as to how auditors make nonstatistical sample size decisions. Specifically, this study seeks to identify the factors that are used by auditors in rendering such judgments in substantive tests of details. Additionally, selected demographic, professional, attitudinal, and behavioral characteristics of these auditors will be measured; and it will be determined if a relationship exists between auditors' judgment policies and the selected characteristics.

The chapter is divided into five main sections. First, the research methodology, specifically Brunswik's lens model, is defined and described. Second, within the context of the lens model this study's research questions are detailed. Third, the sample population of auditors is described. Fourth, the development of the research instrument (questionnaire) designed to capture auditors' judgment policies in nonstatistical sample size decisions is discussed. Section four will also define the demographic, professional, attitudinal, and behavioral variables that are measured in the study. Finally, the statistical tests used to analyze the questionnaire results are detailed.

4.1 THE LENS MODEL

Brunswik's lens model is a general procedure designed for examining the unique information processing behavior of raters under conditions of uncertainty. It portrays the rater as judge of an event that cannot be observed directly (e.g., future business failure) through a 'lens' of cues (e.g., accounting ratios) whose relationship to both the event and the judge are uncertain.[122] The interaction between the individual and the environment is described by a number of relationships, including those among the cues, those between the cues and the criterion event, those between the cues and the judge's response, and those between the criterion event and the judge's response.[123] The multivariate relationships between the judge's response, the criterion event, and the informational cues (i.e., factors) are typically expressed in the form of correlation coefficients. Depending on the specific relationship to which it pertains, the correlation coefficient can be an index of rater judgmental accuracy, judgmental agreement, or the optimality of the rater's factor weighting strategy.

---

[122] Joyce, E. J. and R. Libby, "Behavioral Studies of Audit Decision Making," Journal of Accounting Literature (Vol. I, 1982), p. 103.

[123] Ibid, p. 103.

Basically, the lens paradigm emphasizes the construction of linear equations that model the relationship that exists between the rater's responses and the factors that have been provided. Hoffman termed these linear models a paramorphic representation[124] of judgments because he did not feel that the actual cognitive processes involved in making the judgment was a linear weighting of factors, but that such cognitive processes could be modelled by a linear weighting scheme. In this sense, the linear equation represents the captured rating policy of the individual rater. The captured policy is taken to represent an explicit objective description of the way in which the rater combines and weights dimensional information in arriving at overall ratings.[125] The significance of a statistical equation being used as a model of a rater's judgmental policy is underscored by the following:

> The captured rating policy thus is the crux of the entire (lens model) methodology--with proponents arguing that actual rating behavior can be accurately described in this manner. The implicit assumption is that the captured rating policies are construct valid representations of "true" rating

---

[124] Hoffman, P. J., "The Paramorphic Representation of Clinical Judgment," Psychological Bulletin (March 1960), pp. 116-31.

[125] Hobson, Charles J. and Frederick W. Gibson, "Policy Capturing as an Approach to Understanding and Improving Performance Appraisal: A Review of the Literature," Academy of Management Review (Vol. 8, No. 4, 1983), p. 640.

policies.[126]

The basic policy capturing paradigm requires raters to make quantitative evaluations of a number of cases, each of which is defined by one or more factors with each factor having several levels (values). If there is a sufficient number of cases for which the rater must provide judgments, and each case is described by the same set of factors, then a regression equation can be found which describes each rater's idiosyncratic method of combining information. This regression equation represents the judgmental strategy or policy of the rater, with the beta coefficients representing the relative degree of importance that a given factor has in predicting the criterion. A captured rating policy takes the form of the following multiple regression equation:

$$Y_s = b_o + b_1 X_1 + b_2 X_2 + \ldots + b_n X_n + e_i \quad (1.1)$$

where:
Y_s = represents the optimal linear regression of the rater's judgment.
b_o = the Y intercept, which represents the mean of the distribution of Y when $X_1, X_2, X_n = 0$.
b_i = represents each factor's beta coefficient.
X_i = represents the factors in their standardized form.
e_i = the random error term.

---

[126] <u>Ibid</u>, p. 640.

The key aspect of this procedure is that the rater is not requested to explicitly state his subjective assessment of the importance of the factors, but that the importance of a factor is inferred from the objective beta weight that is calculated on the basis of the rater's responses to all the cases. This approach, which is merely a modelling of the relationship that exists between the inputs (i.e., factors) and outputs (i.e., rater's judgments), has been referred to as a black box[127] research strategy.

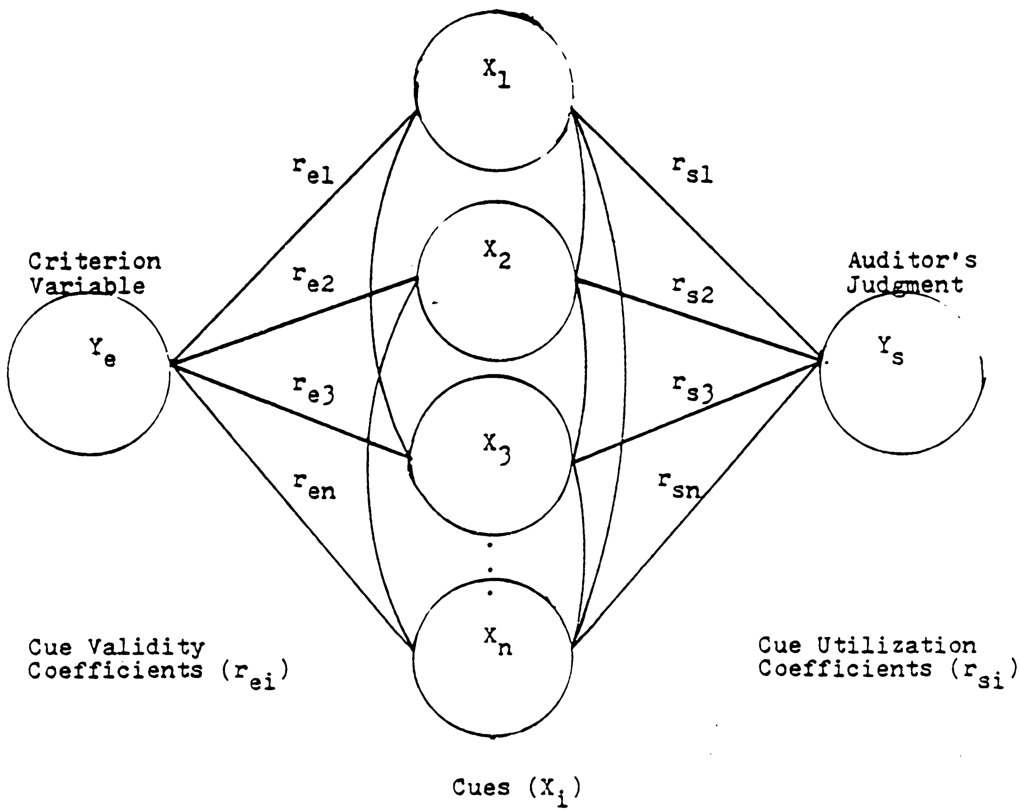
Brunswik's lens model has been described quite thoroughly by Ashton in his recent monograph published by the American Accounting Association.[128] Ashton's description of the lens model (Figure 2, p. 128) is relied on here.

The lens model dissects the universe into two main partitions: (1) the environment, or the left side of the lens; and (2) the rater's judgment scheme, or the right side of the model. There are three basic elements that are contained in the paradigm: (1) the distal variable (Y_e) with which the rater is concerned; (2) the factors (X_i) that may assist the rater in rendering his judgment; and (3) the judgment (Y_s) as actually rendered by the rater.

---

[127] Hayes, J. R., "Strategies in Judgmental Research," in B. Kleinmuntz, ed., Formal Representation of Human Judgment (New York: John Wiley & Sons, Inc., 1968).

[128] Ashton, Robert H., Human Information Processing in Accounting, Studies in Accounting Research No. 17.



* Ashton, Robert H., "Studies in Accounting Research #17," Human Information Processing in Accounting, (American Accounting Association: Sarasota, Florida, 1982), p. 15.

FIGURE 2

Brunswik's Lens Model

In the context of the present study: (1) the distal variable (Y_e), is each case's optimal sample size as determined by the 'mean estimation' statistical sampling technique; (2) the factors are such items as tolerable error and risk of incorrect acceptance, that typically vary across the cases; and (3) the auditor's nonstatistical sample size decision is the variable (Y_s). The Y_e and Y_s values will differ if the statistical relationship between Y_e and the factors is less than perfect, when the rater's use of the factors is sub-optimal relative to the environment, or when the rater has been inconsistent in the exercise of his judgment policy.[129]

Correlation coefficients are employed most often to reflect the relationship between the distal variable and the factors. On the left side of the lens model, these coefficients (r_{ei}), are referred to as validity coefficients. Similarly, where the association is between the rater's judgment and the factors, the coefficients (r_{si}) are referred to as utilization coefficients. The values associated with the utilization coefficients represent the relative degree of importance that the rater implicitly attaches to the factors in rendering his judgmental response. The possible collinearity among the factors is represented by the lines con-

---

[129] Ibid, p. 15.

necting the factor set. Finally, the judgmental accuracy of the rater is depicted by the correlation (r_a) of repeated occurrences of Y_e and Y_s .

4.1.1 Indices of Judgment

The two focal areas of judgment modeling research are: (1) the accuracy of judgments, and (2) agreement (or consensus) among the judgments of raters. The methodology provided by the lens model is well-suited for the assessment of judgmental accuracy and performance. A framework which can be used to assess judgmental accuracy and performance is provided in a modified version of Ashton's, "Indices of Judgment"[130] (Figure 3, p. 131). Ashton's explanations of the variables and indices are relied upon in this section.

---

[130] Ibid, p. 17.

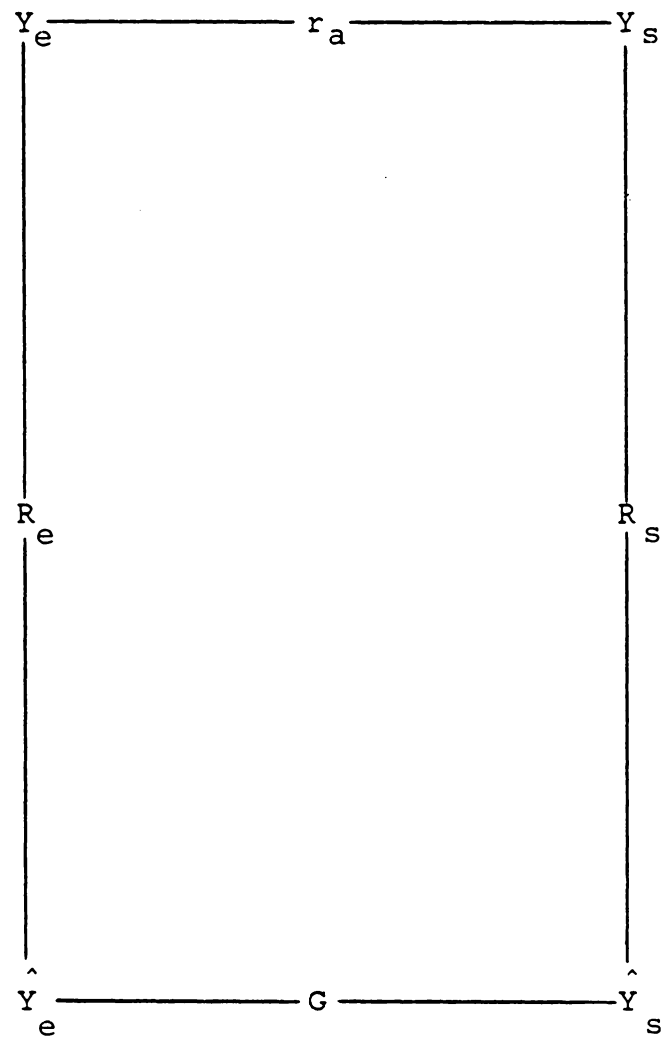


FIGURE 3

Indices of Judgment

A review of Figure 3 will be done within the context of the present study. As defined earlier in this chapter, the variable Y_e represents the observed value, the statistical sample size, that is generated by the 'mean estimation' technique. The auditor's judgment or recommended nonstatistical sample size is the variable, Y_s. The variable, $\hat{Y}_e$, is the optimal linear regression prediction of the 'mean estimation' sample size, Y_e. The index r_a (= $rY_e Y_s$) reflects the auditor's accuracy in predicting the distal value. The $\hat{Y}_s$ is the optimal linear regression prediction of the auditor's judgment. The variable, $\hat{Y}_s$ is formulated by initially regressing the auditor's nonstatistical sample size judgments over a series of cases onto the independent factors. The resultant regression equation that represents the auditor's judgment policy, is then applied to the standardized factor levels to yield the auditor's predicted nonstatistical sample size, $\hat{Y}_s$.

The multiple correlation coefficient R_e is the correlation between the observed (Y_e) and the predicted ($\hat{Y}_e$) 'mean estimation' sample sizes. It represents the extent to which Y_e can be predicted by a linear combination of the independent factors. The squared multiple correlation coefficient, R^2_e, reflects the degree to which variance in the observed statistical sample size (Y_e) can be accounted for by the linear regression equation.

The multiple correlation coefficient R_s is considered a measure of the auditor's consistency, or the extent to which the auditor consistently utilizes his judgment policy as represented by the multiple regression equation.[131] The square of this multiple correlation (R_s^2) indicates the degree to which variance in Y_s is explained by the linear regression model.

The index G ($= r\hat{Y}_e \hat{Y}_s$), represents the degree to which the auditor's predicted judgment ($\hat{Y}_s$) matches the predicted 'mean estimation' value ($\hat{Y}_e$) of the linear model. The matching index is a measure of the extent to which auditors detect the essential properties (e.g., tolerable error, population characteristics) of the judgment task.

Having defined Brunswik's lens model and the corresponding 'Indices of Judgment', a review of how this specific methodology will help evaluate the research questions that have been posed in this exploratory study is appropriate.

---

[131] Ibid, p. 17.

4.2 RESEARCH QUESTIONS

The present study involved both group and individual analyses. The group segment involved the random assignment of each of seventy practicing auditors, chosen from the population of practitioners who primarily audit small businesses, to one cell in a 7X10 matrix. Each matrix cell contains five audit sampling cases for which the auditor was required to render nonstatistical sample size judgments. A total of fifty audit cases were constructed for this study. A random permutation of the fifty cases within each of the seven columns insured that each case was analyzed once by seven different auditors.

To infer which factors are used by the typical auditor in making nonstatistical sample size decisions (research question 1), the average sample size recommendations (Y_s) for each of the fifty cases will be regressed onto the independent factors. The resulting multiple regression equation (see equation 1.1, p.126), represents the captured rating policy of the typical auditor. In this context, reference is made to the typical auditor because there is no compelling reason to believe that the sample selected, is not representative of the population of auditors who primarily audit small businesses and who primarily perform nonstatistical sampling.

In addition to ascertaining how the typical auditor renders nonstatistical sample size decisions, a related question focuses on whether the typical auditor is making optimal or ideal sample size decisions relative to the 'mean estimation' statistical model. The specific methodology employed in evaluating this question will involve a t-test to test the null hypothesis that the average difference between the cases' mean recommended sample sizes and the 'mean estimation' sample sizes is equal to zero. If the typical auditor is properly considering the SAS No. 39 factors the average difference should not be significantly different from zero. Should the t-test lead to a rejection of the hypothesis, then the regression procedure should provide evidence as to which factors best explain the significant finding.

In addition to the group analysis, a sample of eight auditors who are highly representative of the group of seventy was selected. Each of these eight auditors was asked to make nonstatistical sample size decisions for each of the fifty cases that were used in the group analysis. The captured rating strategy of each is inferred by regressing the auditor's fifty sample size decisions onto the independent factors. Thus, the individual analysis produces eight captured rating policies, with each being similar in form to equation 1.1.

The group analysis provides evidence as to how the typical auditor is making nonstatistical sample size decisions in substantive tests of details. As such, it identifies the factors that are implicitly used by the typical auditor in nonstatistical sampling applications. The inherent advantages of the group analysis are two-fold: (1) since each auditor is only analyzing five cases, systematic effects attributable to fatigue, boredom, etc., are minimized, and (2) from a practical perspective, the small case load for each auditor permitted the selection of a larger sample of practitioners. However, should the group analysis reveal that the typical auditor properly considers the factors that SAS No. 39 describes, there is no way of judging from that analysis alone whether specific auditors are in conformity with the Statement. That is, with each of the seventy auditors not having analyzed the same five cases, it will not be possible to test whether significant differences exist between the auditors' judgment policies. It is because of this limitation in the group design that the individual analysis of the eight auditors was performed.

The group analysis permits an evaluation of whether the audit profession as a whole, as represented by those auditors who primarily apply nonstatistical sampling methods, is in conformity with SAS No. 39. However, it tells us little with respect to specific, individual auditors. The analysis of the eight in a completely crossed factorial design will serve to strengthen and complement the group analysis. A two-way analysis of variance (ANOVA) model will be used to evaluate whether significant differences exist between auditors' judgment strategies.

4.2.1 Judgmental Accuracy

Research question 2 (How accurate are the nonstatistical sample size recommendations of the auditor when compared to the 'mean estimation' statistical sample sizes?) can best be evaluated by aggregating the sample size predictions on the fifty cases that were used in both the group and individual analyses. This design will produce a total of fifteen observations for each of the fifty cases, or a total of 750 observations (i.e., recommended sample sizes, Y_s). Using the 'mean estimation' statistical model (which explicitly takes into consideration some of the factors that SAS No. 39 defines), the optimal sample size (Y_e) for each of the cases will be determined. The Pearson product moment correlation ($r_a = rY_e Y_s$) between the auditors' recommended sample sizes and the statistical model's sample sizes will provide evidence as to the accuracy of auditors' judgments.

4.2.2 Judgmental Agreement

Before focusing on the topic of judgmental agreement, it is appropriate to explain the importance of judgmental agreement in audit decision making. Although judgmental accuracy will be evaluated in this study, there are many audit judgment situations where it is impossible to assess the accuracy of auditors' decisions. For example, since no definitive criteria presently exist for defining the optimum internal control system, auditor evaluations of internal control cannot be evaluated on the accuracy dimension. Similarly, the absence of objective criteria for determining what constitutes materiality, precludes the assessment of accuracy on this particular issue.

The inability of researchers to measure judgmental accuracy in many audit issues has not lessened the audit profession's concern for the aspect of auditor judgment. In part, this is because when auditor judgments are questioned in litigation or regulatory proceedings, successful defense often entails establishing a professional consensus (via expert witnesses) that the defendent acted in a prudent manner.[132] In light of this explanation, one of the major goals of the AICPA's continuing professional educational program

---

[132] Joyce, Edward J., and Robert Libby, "Behavioral Studies of Audit Decision Making," Journal of Accounting Literature (1982), p. 105.

is to enhance agreement among auditors' decisions. Similarly, many large public accounting firms promote auditor agreement by using extensive worksheets and detailed procedure manuals. For example, one large multi-national public accounting firm that participated in this study's pilot test, has developed detailed worksheets to promote judgment agreement among its auditors' nonstatistical sample size decisions.

Research questions 3 and 4 address the other major area of interest that has been guided by the lens model, judgmental agreement. Judgmental agreement may be considered within two contexts: (1) agreement in fact, and (2) agreement in principle. The former concerns agreement with respect to overall judgments, while the latter concerns agreement with respect to the captured policies generating the overall judgments.[133]

The most popular type of agreement that has been assessed in judgment modeling is consensus, or the collective agreement across raters using identical data at the same point in time. Research question 3 (What is the level of agreement (i.e., consensus) across auditors' nonstatistical sample size judgments?) will be evaluated in this study by the intraclass correlation. The intraclass correlation

---

[133] Ashton, Robert H., Human Information Processing in Accounting, Studies in Accounting Research No. 17, p. 45.

evaluates the agreement or consistency in patterns of recommended sample sizes between the auditors' judgments across the fifty cases.

Correlational analysis will be used in this study to evaluate research question 4 (How similar are different auditors' judgment policies for making nonstatistical sample size decisions?). Specifically, for each of the eight auditors who evaluated all fifty cases the auditors' predicted sample sizes ($\hat{Y}_s$) will be determined. The degree of association among the auditors' predicted judgments will be evaluated by correlating the responses of all subject pairs and taking the average.

4.2.3 Judgmental Insight

The fifth and final research question focuses on auditor self-insight (How much self-insight do auditors exhibit in their judgment policies?). This phase of the study can be accomplished by initially having each of the eight auditors who analyzed all fifty cases, allocate 100 points across the factors (subjective weights) to indicate the relative importance the auditor feels he has placed on each factor. This subjective weighting scheme was requested after the auditor rendered his judgments. A measure of the degree of auditor self-insight can be obtained by computing the mean association between the ranks of the standardized beta weights as derived from the regression procedure and the ranks of the auditors' subjectively assigned weights. The Spearman rank correlation procedure will be used to assess the degree of association.

4.3 THE POPULATION

Since the principal objective of this study was to determine how auditors make nonstatistical sample size decisions, it was imperative that a sample of subjects be selected who routinely apply nonstatistical sampling techniques. Such a sample could be expected to come from practitioners associated with local or regional CPA firms, who primarily audit small businesses. Accordingly, it was decided to exclude from this study the large, multi-national public accounting firms that almost exclusively rely on statistical sampling approaches.

As noted previously in this paper, the specific research methodology of this study involved both a group and individual analyses. The group analysis required each of seventy auditors, who are representative of the typical small practitioner, to make nonstatistical sample size decisions for each of five fictitious audit cases. A total of fifty cases were prepared for this study. In the individual analysis, each of eight auditors, who are not unlike the group of seventy in professional and demographic characteristics, rendered sample size decisions for each of the fifty cases that were used in the group analysis.

All of the auditors who took part in the experiment were associated with public accounting firms in various cities in the state of Virginia. The participants were not randomly selected, but were chosen on the basis of previous knowledge regarding their desire to participate. Thus, this research could be limited by the restrictive nature of the sample which may preclude the generalization of the study's findings to the population of auditors as a whole.

Of the seventy practicing auditors who participated in the group analysis, sixty-four (91 percent) were certified public accountants. The auditors were employed by twenty-seven public accounting firms that were located in Hampton, Harrisonburg, Newport News, Norfolk, Richmond, Roanoke, and Winchester, Virginia. On the average, each of the group auditors had more than 8 years of auditing experience. The median experience level was 6 years, while the range of professional expertise was from 18 months to 34 years. Based on consultation with auditing personnel, it was ascertained that the typical auditor, employed by the typical auditing firm, can be expected to routinely render sample size judgments after four years of experience. When asked, "What proportion of your total sampling could be characterized as nonstatistical?", thirty-seven auditors replied "100 percent." On the average, the respondents indicated that 81 percent of their testing applications could be characterized as nonstatistical.

Eight practicing auditors, all of whom were certified, participated in the individual analysis. They were employed by seven different public accounting firms with audit practices in Alexandria, Charlottesville, Fairfax, Hampton, Harrisonburg, Newport News, and Winchester, Virginia. The respondents possessed an average of 7.04 years of experience, with only one auditor having less than five years experience (four years). The auditor with the most extensive auditing background, had twelve years of auditing expertise. The median level of experience for the participants was 6.25 years. On average, the respondents classified 97 percent of their sampling applications as nonstatistical. The median proportion was 99.5 percent.

The group study was conducted by an interview process in which the author visited each of the twenty-seven audit firms over the course of a twelve-day period, from November 7 to November 18, 1983. Typically, the experiment was conducted in the privacy of the firm's conference room or reference library.

Prior to commencing the case analysis, the auditors were given an oral explanation of the judgment task. Furthermore, each respondent's anonymity was assured prior to participating in the experiment. For the duration of the experiment the author remained with the auditors in the event any questions were posed. On average, the auditors required 30 minutes to complete the judgment task with only a small minority requiring more than forty minutes to conclude the exercise.

Because of the time estimate (as determined from a pilot test) required to analyze all fifty cases, it was not possible to conduct the individual analysis by interview. Instead, each of the eight auditors at his own discretion, spent five to eight hours completing the judgment task. Before starting the experiment, each of the auditors was urged to refrain from analyzing the fifty cases in one sitting. This appeal was made in order to minimize any systematic effects due to fatigue, boredom, etc. It was recommended that the cases be analyzed in non-continuous, one-hour segments.

Finally, it is strongly believed that the vast majority of auditors in the study were willing, enthusiastic, and diligent participants. The majority of firms and auditors were initially identified in the Virginia Society of Certified Public Accountants, Membership Directory. A screening was conducted by telephone to insure that each firm and/or auditor matched the desired profile characteristics. During the course of the experiment, some individuals were disqualified when it was apparent that the study's objectives would be compromised by their participation. Specifically, five auditors were disqualified because of their inexperience in rendering nonstatistical sample size judgments.

4.4 DEVELOPMENT OF THE QUESTIONNAIRE

To evaluate this study's research questions, it was necessary to create an instrument that would properly elicit auditors' nonstatistical sample size judgments. In pursuit of this objective, it was necessary for the author to understand completely the terms, definitions, and sampling concepts that SAS No. 39 enumerates. Furthermore, the technical guidance that is provided in the related Audit Sampling Guide needed to be fully comprehended.

The narrative format for all the cases was identical. In each case the auditor was responsible for the design of a nonstatistical sample to test the December 31, 19X0 accounts receivable of the fictitious electrical supply company, ABC Incorporated. The decision not to vary the case scenario was made in the hope of minimizing the time that each participant would need to become acquainted with a case's background information. Accordingly, it was felt that the auditors would be able to devote more of their time to the required judgment task.

The cases' narrative was taken primarily from the non-statistical sampling case study that was detailed in the Audit Sampling Guide (pp.64-5). In accordance with the Guide's example, the cases' accounts receivable were highly stratified to facilitate the auditor's decision process. Each case described in detail, from the total number of accounts receivable: a) those balances that would be examined 100 percent (and thus, should be excluded from the population to be sampled), b) those credit balances that would be tested separately as accounts payable, c) those balances of minimal value that should be ignored, and d) those specific accounts receivable from which the auditor was to determine the appropriate sample size.

For the most part, each of the factors that should be implicitly considered by the auditor in his judgment decisions, were highlighted separately in the lower half of the questionnaire. For example, the auditor's tolerable error and alpha risk were each identified and represented in scaled graphic form (see figure 4, p.148). The presentation of the factors in this format was expected to insure clarity and ease of evaluation. The factor identified with variation (i.e., standard deviation) in the specific group of accounts from which the auditor was to select a sample size, was included in the narrative segment of the cases. Similarly, the factors that pertained to last year's sample size (Past), the firm's sales (Sales), and the dollar value of the accounts receivable (Asset) were included in the cases' narrative segment.

The factor, risk of incorrect acceptance, was not presented per se as a quantification level. Instead, the factors that influence the auditor's acceptable level of risk of incorrect acceptance were presented. The factors that are used to assist the auditor in planning his risk of incorrect acceptance (TD) are: (1) ultimate risk (UR), (2) the risk of undetected error due to internal control failure (IC), and (3) the risk of failing to detect errors by other substantive tests directed toward the same specific audit objective (AR).[134]

---

[134] Statistical Sampling Subcommittee, Audit Sampling Guide, Appendix F, p. 123.

Case No. L9807

You are to assume that you are responsible for the design of a nonstatistical sample to test the December 31. 19X0 accounts receivable of the ABC Company. ABC Inc., an electrical supply company, has been a client of your firm for the past several years.

For the year ended December 31, 19X0, ABC had sales of approximately $ 2 million. At December 31, 19X0, the population contained 500 accounts receivable, with debit balances aggregating $ 350,000 . These balances ranged from $ 1 to $ 10,000. There were also 6 credit balances aggregating $ 2100, which you have decided to test separately as accounts payable. Total assets on the year-end balance sheet were $ 1,200,000.

The population contains eight balances over $5000, which total $55,000. You have decided to examine these balances 100 percent and exclude them from the population to be sampled. The population also contains 150 balances equal to or greater than $ 400 which total $ 250,000 . It is from this specific group of 150 that you are to determine the appropriate sample size. You have estimated that the standard deviation of this group is $ 185 . Last year the sample size which you selected was 72

Assume that you have made the following judgments and estimates which may or may not be useful to you in your sample size decision.

Your assumed level of ultimate risk that any existing monetary errors greater than tolerable error might remain undetected in the account balance after you have completed all audit procedures deemed necessary is:

1 out of 100 — 1 out of 20 — 1 out of 10 — 1 out of 5

Based on your preliminary estimate of materiality levels, your subjective assessment of tolerable error in the 150 customer account receivable balances that might result in a material misstatement of the financial statements is:

$ 20,000. — $ 30,000 — $ 50,000. — $ 60,000.

Your subjective assessment of risk that internal control might fail to detect errors greater than the tolerable error is:

50% — 60 % — 70 % — 100%

Your subjective assessment of risk that analytical review procedures and other substantive tests might fail to detect errors greater than the tolerable error:

40% — 50 % — 60 % — 100%

Your subjective assessment of the dollar value of errors in the population is:

$ 5,000. — $ 6,000. — $ 8,000. — $10,000.

The risk (alpha) that the sample erroneously supports the conclusion that the account balance is materially misstated when in fact it is not:

1 out of 100 — 1 out of 20 — 1 out of 10 — 1 out of 5

After considering the above information, what sample size (without replacement) do you recommend for the above case? ________________.

FIGURE 4

Case Example

The auditor should arrive at an appropriate risk of incorrect acceptance by relating the above factors in the following mathematical model: TD = UR/(IC X AR). In a realistic auditing situation, the auditor exercises professional judgment in specifying an acceptable ultimate risk (UR) and subjectively quantifies his judgment of the risks 'IC' and 'AR.'[135] However, in the experimental cases it was not necessary for the participants to quantify the factors, since all of the factor levels had been pre-established by the author.

In designing the fifty cases, it was necessary that the factor levels be highly reflective of an audit situation where nonstatistical sampling is appropriate. For example, in the vast majority of experimental cases, the risk that internal control (IC) might fail to detect material errors ranged from .80 to 1.00. This is because the small business environment (where much nonstatistical sampling is applied), is often lacking in adequate internal accounting controls. In such situations, auditors are precluded from performing much compliance testing and typically adopt a substantive approach to their audit.

---

[135] Ibid, p. 123.

The objective in developing the cases' factor levels was to insure that each case possessed the highest degree of external validity. A matrix of the cases' factor levels, 'mean estimation' statistical sample sizes, and the auditors' average recommended nonstatistical sample sizes is given in Appendix A. For each factor, general guidelines were instituted after consultation with representatives from an international public accounting firm and this author's primary auditing advisor. The following guidelines were observed in constructing the factor levels in the vast majority of cases. Consistent with a substantive auditing approach, the internal control risks were typically .80 or 1.00. Similarly, the risks identified with analytical review procedures ranged from 50 to 100 percent. The value of the factor, tolerable error (TE), generally approximated 5 to 8 percent of stratum book value. The value of the factor, expectation of error (E), was typically equal to 30 percent of the cases' tolerable error. In all but a few cases, the factor, ultimate risk, was either .05 or .01. The factor, risk of incorrect rejection (alpha), in most cases was either .05 or .01. Finally, the number of accounts (N) in the particular stratum from which the auditor was to select a sample, generally ranged from 80 to 600.

Table 1 (p. 152) lists all of the independent factors that have previously been discussed. For each variable, the mean, minimum, and maximum values are given. In addition to the factors that are defined in the Audit Sampling Guide, last year's sample size (Past) was also included in each case as a possible influencing factor. This was done to ascertain its influence (if any) in the typical auditor's judgment scheme. Based on the results of the pilot study, two additional independent variables ('Asset' and 'Sales') were included in each case as possible explanatory factors. The values for the criterion variable, 'mean estimation' statistical sample size, are also reflected in the table.

TABLE 1

Independent Factors' Values

Variable	Mean	Minimium	Maximum
Sigma	$127.	$50.	$250.
UR	.049	.01	.20
AR	.657	.50	1.00
IC	.804	.20	1.00
TD	.114	.01	.33
TE	$16,830.	$3,000.	$50,000.
E	$4,932.	$1,500.	$15,000.
N	221	80	600
Alpha	.042	.01	.10
Past	32	0	87
Asset	$211,700.	$50,000.	$450,000.
Sales	$2,140,000.	$1M.	$3M.
Criterion	29.54	7	97

4.4.1 Demographic, Attitudinal, and Professional Characteristics

The demographic, attitudinal, professional, and behavioral characteristics that were measured were chosen because these variables have intuitive merit as those that might predict and explain (accuracy or inaccuracy) in estimating required sample sizes. Furthermore, the selection was made with the belief that information about the relationship between these characteristics and auditors' nonstatistical sample size decisions will provide a basis for understanding auditors' judgment policies relative to such decisions.

Each packet of cases that was distributed to the experiment's respondents contained two additional sections. One section (see Figure 5, p. 154) consisted of questions that measured the demographic, professional, behavioral, and attitudinal variables that have just been listed. The remaining section (to be discussed later in this chapter), was included to determine the degree of auditor self-insight.

Questionnaire

Instructions: Would you kindly take a few additional minutes to respond to the following questions? You may be assured this information will be kept in strict confidence. The results will be used only in statistical summaries; no respondents will be individually identified.

1. Age? ______Years 2. Sex? (circle) M F

3. What is the highest level of formal education you have completed? (circle)
   a. High school
   b. High School plus some college
   c. Bachelors Degree (Institution?_______________)(Major?________)
   d. Masters Degree (Institution?_____________)(Major?_____________)
   e. Law Degree (Institution?______________) (Degree?_____________)
   f. Other (Explain____)

4. Are you a CPA? (circle) Yes No. If Yes, when did you become certified? ________Year, ________Month.

5. How many courses have you had in statistics? ________Number.

6. How long have you been associated with the present firm? ____Years.

7. What is your primary area of responsibility in this firm? (circle)
   a. Auditing b. Taxation c. MAS d. Other ____________

8. How much auditing experience do you have? ________Years _______Mos.

9. What is the average length of engagement for your typical audit client? ___________Months.

10. What percentage of your total sampling could be characterized as nonstatistical (judgmental) ? _________%

11. Have you taken any CPE courses or seminars on the topic of "audit sampling"? (circle) Yes No. If Yes, when? _______Yr. _____Mo.

12. How well do you understand SAS No. 39, "Audit Sampling"? (circle)
   a. Complete understanding
   b. Adequate understanding
   c. Insufficient understanding
   d. No understanding

13. What effect has SAS No. 39 had on your nonstatistical (judgmental) sampling applications? (circle)
   a. Strong effect b. Moderate effect c. Little or no effect

14. How would you describe your general reaction to SAS No. 39? (circle the appropriate letters)
   a. The Statement provides needed guidance and direction.
   b. The Statement is too technical for my judgmental sampling tasks.
   c. The Statement unduly influences me to adopt statistical sampling.
   d. The Statement is a classic example of standards overload.
   e. Other (explain____.)

FIGURE 5
Demographic and Professional Characteristics

4.4.2 Self-Insight Questionnaire

As indicated earlier in this chapter, the experiment required each of the respondents to allocate 100 points across the factors as an indication of how he perceived each factor's relative influence in the decision task. Although this elicitation technique is quite simple, it has been shown to yield factor weights that reproduce actual predictions as well as those obtained by more complex elicitation methods.[136]

The majority of the studies that have been performed in the area of audit decision making, have assessed auditor self-insight by computing the correlations between subjective weights and standardized regression weights. Accordingly, this study will also assess auditor self-insight by this procedure. The evidence provided by this specific statistical methodology indicates that: (1) the degree of association between subjective weights and statistical weights generally is very low, (2) individuals typically overweight relatively unimportant factors, and (3) individuals typically think they utilize more factors than is revealed by the statistical weighting technique employed.[137]

---

[136] Ashton, Robert H., Human Information Processing in Accounting, Studies in Accounting Research No. 17, p. 40

[137] Ibid, p. 40.

The instrument used to record each auditor's subjective weighting policy is shown in Figure 6 (p.157). During the course of conducting the experiment, many auditors were observed to allocate some portion of the 100 allotted points to each of the factors that have been described.

Taking into consideration both the group and individual analyses, a total of seventy-eight auditors participated in this study. Each of the respondents answered nearly all of the inquiries on the questionnaire that measured demographic, professional, behavioral, and attitudinal characteristics. A 100 percent response rate was found on the self-insight questionnaire.

Dear Respondent:

The study which you are participating in is classified within the broad category of "human information processing." It is common in a study of this type to assess the participant's self-insight into which factors he feels are influential in his judgmental process.

An index of self-insight is derived by initially asking you to allocate 100 points across the factors which are listed below. Your basis for allocation is the relative importance that you feel you have placed on the factors in rendering your judgmental sample size decisions. Of course, if there are factors listed below which have been ignored by you in the judgmental task, then you should properly allocate 0 points to these factors.

Factor Description	Points
Ultimate risk level	________
Materiality	________
Internal control risk	________
Analytical review risk	________
Expectation of error	________
Alpha risk (incorrect rejection)	________
Sales $$$$	________
Total assets $$$$	________
Number of accounts	________
Standard deviation	________
Last year's sample size	________
Other (define) ________	________
Total Points Allocated	100

FIGURE 6
Self-Insight Questionnaire

4.4.3 Pilot Study

Prior to the actual experiment being conducted, a sample of the hypothetical cases were administered to representatives of a large, international public accounting firm. The firm selected was chosen because it had been instrumental in developing SAS No. 39.

The pilot study was designed to answer several important questions regarding the fictitious cases--questions that primarily focused on the cases' internal and external validity. For example, were the relationships between the factors (e.g., sales, total assets) logically consistent? Furthermore, were the values that had been established for the factors representative of a situation where nonstatistical sampling was appropriate?

Upon reviewing and working through a sample of the cases, the audit firm was of the general opinion that the cases were internally and externally valid. However, based on the suggestions that were recommended, two modifications were made to the cases. First, the final version of the cases contained values for the variables, sales and total assets. Figures for these items had not been included in the cases' initial draft. Second, the factor, expectation of error, as a percentage of the factor tolerable error, was amended so that the former averaged 30 percent of the latter factor's value. Prior to the change, expectation of error was typically 40 percent of tolerable error. It was the firm's opinion, that recurring audit situations where the auditor's expectation of error exceeded one-third his estimate of tolerable error were highly unusual.

In addition to those suggestions, the firm was helpful in indicating the general guidelines that it observes with some of the cases' factors. For example, as a matter of firm policy, the auditor's risk of incorrect rejection (alpha) is set at 5 percent. In this particular study, the alpha risk averaged .0424 across the 50 cases, with all values set at either .10, .05 or .01. Furthermore, it was the firm's recollection that tolerable error (the firm's nomenclature was 'gauge'), typically approximated 5 percent of stratum book value. In the experimental cases, tolerable error as a percentage of stratum book value had a mean value of 7.6 percent.

To obtain an estimate of the time required to complete the decision task, two doctoral students in accounting at Virginia Tech, rendered their individual judgments on a sample of five cases. Both students who participated in this segment of the pre-test, had extensive public accounting experience prior to commencing their doctoral studies. In each situation, the students took less than thirty minutes to complete the judgment task.

The final phase of the pre-test involved this author analyzing two unique groups of twenty-five cases each. In one three hour session, the author rendered nonstatistical sample size decisions for half of the cases. The remaining group of cases were analyzed over the course of three one-hour, non-continuous time segments. For the group of cases, which were analyzed over three one-hour time periods, the judgmental accuracy coefficient was 0.531. Conversely, the index of accuracy on the other group of cases was .37. These findings were used to support the recommendation that the auditors in the individual analysis refrain from analyzing all of the cases in one sitting.

4.5 STATISTICAL ANALYSIS EMPLOYED

The statistical analysis had five goals:

1. to describe the judgment policy or strategy of the typical auditor relative to nonstatistical sample size decisions;

2. to determine whether the nonstatisctical sample size decisions of the typical auditor correspond to the 'mean estimation' statistical sample sizes.

3. to ascertain whether there are significant differences in the judgment policies of specific auditors;

4. to evaluate the extent of auditors' judgmental accuracy, agreement, and self-insight; and

5. to determine the relationships that exist between the independent variables of demographic, professional, attitudinal, and behavioral characteristics and the judgmental accuracy, agreement, and self-insight of auditors.

To attain these objectives, five categories of statistical analysis were employed.

Numerous descriptive statistics were used to describe the relationships contained in the data. It is assumed that all of these statistical measures are well understood by researchers in most academic fields and, thus, they will not be detailed here. Means, ranges, and standard deviations were computed for auditors' nonstatistical sample size decisions. The following sections describe the statistical methods that comprised this study's methodology.

4.5.1 Multiple Regression

There are two basic approaches that are available for capturing a rater's judgment policy: multiple regression, and analysis of variance (ANOVA). Of the two, multiple regression is decidedly more popular, although each technique has its acknowledged benefits. Multiple regression has been selected as this study's methodology to capture the typical auditor's judgment scheme.

As indicated earlier in this chapter, the typical auditor's judgment policy is found by regressing mean nonstatistical sample size decisions onto the independent factors. The resulting linear equation models the typical auditor's actual cognitive process. The relative importance of each factor in predicting the criterion variable is inferred from the regression coefficients that are calculated on the basis of the auditors' judgments over all the cases. An example of a linear multiple regression equation which is representative of the typical auditor's judgment policy is given in equation 1.1 (p.126).

Equation 1.1 implies that auditors combine information in a linear additive way. That is, the effect on the criterion variable of a change in one predictor factor is independent of all other factor levels. In addition to linear models, it is also possible to have models which incorporate configural and/or curvilinear terms. Configural models are represented by the inclusion of cross-product or interaction terms (e.g., $b_{sij} X_i X_j$). The following example reflects the configural use of factors:

> Outstanding strength or weakness can have precisely opposite meanings at different times in the market cycle. For example, consistent strength and volume in a particular issue, occurring after a long decline, will usually turn out to be an extremely bullish indicator.... On the other hand, after an extensive advance which finally spreads to issues neglected,...individual strength and volume are not likely to be short lived, but may suggest the end."[138]

Curvilinear models which are represented by the inclusion of exponential terms (e.g., $b_{si}X_i^2$), are fitting where the relationships beteen the factors and the dependent variables are not linear. For example, the personnel officer who evaluates applicants for the sales position may feel that applicants who score in the middle range of an intelligence test are more likely to be successful than those who score at either extreme.[139]

There are many individuals who claim that the linear additive model is inappropriate to capture their cognitive processes. Such individuals argue strongly for judgment models that include configural and/or curvilinear terms. On the surface, these arguments have a strong practical appeal. For example, a physician may believe that body temperature is related to the probability that a patient has a particular disease if the patient also has symptom S, but that body

---

[138] Loeb, G., The Battle for Investment Survival (New York: Simon & Schuster, 1965), p. 65.

[139] Ashton, Robert H., Human Information Processing in Accounting, Studies in Accounting Research No. 17, p. 19.

temperature is irrelevant to the diagnosis if the patient does not have sympton S.[140]

The use of at least one configural term in the auditor's judgment strategy seems appropriate for this study's sample size judgment task. The auditor's allowable risk of incorrect acceptance (TD), is related to the level of ultimate risk (UR), and the risks which the auditor perceives in internal control (IC) and analytical review(AR). That is, TD = UR/(IC X AR). This being the case, the judgment strategy of Equation 1.1 could be amended to incorporate configurality by the inclusion of some cross-product terms.

Despite the intuitive appeal for using non-linear equations to model raters' judgment policies, a plethora of research findings strongly support the robustness of linear models.[141] That is, studies comparing the relationship between actual judgments (Y_s) and the predicted judgments ($\hat{Y}_s$) generated by linear and nonlinear models, have found only minimal increases in predictive power (i.e., $R_a = rY_eY_s$) as a result of non-linear terms.[142] For example, Wiggins and Hoffman evaluated the increase in predictive power resulting from adding non-linear terms to raters' linear models. For

[140] Ibid, p. 19.

[141] See Slovic & Lichtenstein (1971) for a review.

[142] Ashton, Robert H., Human Information Processing in Accounting, Studies in Accounting Research No. 17, p. 20.

an extensive non-linear model that included 77 terms (11 linear, 11 curvilinear, and 55 configural), the greatest increase in predictive power that resulted from adding these terms was a mere four percent.[143]

Dawes and Corrigan offer four reasons to explain the robustness of linear models over deviations from linearity: (1) linear models are good approximations of any multivariate model that is conditionally monotonic in each independent factor; (2) relative weights in linear regression models are impervious to error in the dependent variable; (3) error in the measurement of independent factors makes optimal functions more linear, and (4) deviations of weights from their optimal values do not make great differences in the resulting composites.[144] Goldberg believed that although "human judges behave in fact in a rather configural fashion, the power of the linear regression model is so great that it serves to obscure the real configural processes in judgment."[145]

---

[143] Wiggins, N., and P. Hoffman, "Three Models of Clinical Judgment", Journal of Abnormal Psychology, (February 1968), pp. 70-77.

[144] Dawes, R. M., and B. Corrigan, "Linear Models in Decision Making," Psychological Bulletin (February 1974), p. 99.

[145] Goldberg, L. R., "Simple Models or Simple Processes? Some Research on Clinical Judgments", American Psychologist (July 1968), p. 488.

The assumptions underlying the multiple linear regression model are:

(1) Each error term is normally distributed with zero mean and constant variance (σ).

(2) The error components are independent across observations.

(3) B_1, B_2, B_3, ..., B_k are (K + 1) parameters, and X_1, X_2, ...X_n are known constants.[146]

4.5.2 Analysis of Variance

Because each of the eight auditors in the individual analysis analyzed the same fifty cases, analysis of variance will be used to determine whether there are significant differences in the judgment policies of specific, individual auditors. Specifically, a two-way analysis of variance is a technique used to test simultaneously the significance of the difference among several means, where two bases of classification (i.e., auditors and cases) exist. Accordingly in this study, the method will be used to test the following null hypotheses: (1) the differences in the mean nonstatis-

[146] Pfaffenberger, Roger C., and James H. Patterson, Statistical Methods for Business and Economics (Homewood, Illinois: Richard D. Irwin, Inc., 1981), p. 484.

tical sample size judgments of all auditors are zero. The appropriate alternative hypothesis to the above, is that the difference between any two of the mean test scores is not zero, (2) the pattern of sample size judgments across the fifty cases is identical for all auditors.

The assumptions underlying the analysis of variance are:

(1) Homogeneity of variance--the population dependent variable variances associated with each treatment in the experiment are equal.
are equal.

(2) Normality--the distribution of dependent variable scores for each treatment population is normal.

(3) Independence--the errors in the treatment populations are statistically independent.

A rejection of hypotheses (1) or (2) would lead one to conclude that auditors are making different sample size decisions. In essence, auditors are applying different procedures or techniques that yield different sample sizes.

4.5.3 Pearson Product Moment Correlation

Pearson Product Moment correlation (r) will be used to measure the extent of auditors' judgmental accuracy and agreement in principle. The reader is referred to the section, 'Indices of Judgment' (pp. 129-131), for a complete explanation as to how the above correlations will be measured in this specific study.

By utilizing the information available in the demographic, professional, behavioral, and attitudinal questionnaire, it will be possible to determine whether relationships exist between these variables and auditors' judgmental accuracy and agreement. For example, it will be determined whether auditors with experience that exceeds the median level, exhibit greater levels of judgmental accuracy and agreement than that attained by their colleagues with experience less than the median amount.

The statistical significance level (p-value) will be determined for each correlation that is computed.

The following two hypotheses will be evaluated in this correlational analysis:

Judgmental Accuracy r_1 : The correlation between auditors' nonstatistical sample size judgments and the 'mean estimation' sample sizes over the fifty cases is not significantly different from zero.

Agreement in Principle — r_2: The average correlation between auditors' predicted judgmental sample size decisions ($\hat{Y}_s$) is not significantly different from zero.

4.5.4 Spearman Coefficient of Rank Correlation

Because normality cannot be assumed with the transformation of the standardized beta coefficients that are used to evaluate auditor self-insight, the Spearman coefficient of rank correlation (R) will be utilized to measure what association exists between the statistically derived regression weights and the auditors' assigned subjective weights. Spearman R is a Pearson correlation between two ordinal (rank order) variables of measurement

The following hypothesis will be evaluated by the Spearman coefficient of rank correlation:

Self-Insight — R : The coefficient of association between auditors' subjective weights and the statistically derived weights is not significantly different from zero.

4.6 CHAPTER SUMMARY

In this chapter, the research methodologies that this empirical study used were reviewed. Brunswik's lens model provides the basic procedure for examining the unique information processing behavior of auditors under conditions of uncertainty. Essentially, the lens paradigm emphasizes the construction of linear equations that model the relationship that exists between the auditor's responses and the factors that have been provided.

Multiple regression was used to describe the judgment policy or strategy of the typical auditor relative to nonstatistical sample size judgments. The typical auditor's judgment policy is determined by regressing mean nonstatistical sample size judgments onto the independent factors. The relative importance of each factor in predicting auditor sample size judgments, is inferred from the beta coefficients that are calculated on the basis of the auditors' judgments over a series of fictitious cases.

Analysis of Variance (ANOVA) was applied to ascertain whether there are significant differences in individual auditors' nonstatistical sample size decisions. Specifically, the procedure was employed to test the significance of differences among the auditors' mean nonstatistical sample size judgments.

The intraclass correlation was used to evaluate the agreement or consistency in patterns of recommended sample sizes between the auditors' judgments across the fifty cases.

Finally, Pearson's Product Moment correlation was utilized to measure the extent of auditors' judgmental accuracy and agreement in principle.

Chapter V

RESULTS OF THE EMPIRICAL STUDY

The research methodology used in the present study was described in detail in the previous chapter. In the present chapter, the statistical results of the empirical examination are reported and discussed.

As discussed in Chapter I, the primary objectives of this study are to determine: (1) how auditors make nonstatistical sample size judgments in substantive tests of details, (2) how their decisions correspond to the 'mean estimation' statistical sample sizes, and (3) whether significant differences exist between different auditors' sample size decisions. Secondary objectives focus on the accuracy and consensus of auditors' sample size decisions. Additionally, the amount of self-insight that auditors exhibit into their judgment policies will be assessed.

Before reviewing the results of the statistical tests, selected auditor characteristics that were measured in the 14-item questionnaire will be discussed.

5.1 AUDITOR CHARACTERISTICS

In this section selected demographic, professional, behavioral and attitudinal characteristics of auditors will be discussed. The primary objective in identifying such characteristics is to ascertain if these characteristics are associated with auditors' nonstatistical sample size judgments. In other words, do such characteristics help explain auditor decisions of this type.

5.1.1 Demographic Characteristics

The first three questions on the questionnaire requested the auditors to provide selected demographic information, including age, sex, and highest educational level.

The average age of the auditors who participated in the study was 32.8 years, with a range from 23 to 58 years. Ten (12.8 percent) of the seventy-eight respondents were female.

Table 2 (p. 174) depicts the highest educational levels attained by the auditors. Only one of the respondents had no college training. Three of the auditors held advanced degrees. The majority of the auditors completed four years of college and majored in accounting. The other areas or disciplines that the auditors majored in include mathematics, economics, and commerce.

TABLE 2

Educational Characteristics

Highest Educational Level	Number	Per Cent
High school	1	1.3
Associate's degree	3	3.8
Bachelor's degree	71	91.1
Master's degree	3	3.8
Total	78	100.0
Academic major		
Accounting	56	71.8
Business Administration	7	9.0
Other	6	7.7
Not classified	9	11.5
Total	78	100.0

5.1.2 Professional and Behavioral Characteristics

The auditor characteristics included in this section were identified primarily because of the expectation that they would help explain the variability among the respondents' sample size decisions. These characteristics are also important when considered alone because they promote our knowledge of the general attributes possessed by auditors.

Seventy-two (92 percent) of the seventy-eight auditors who participated in the experimental study indicated that they were certified public accountants (CPA). When the CPAs were asked to indicate the length of time they have been certified, the average time was seventy-seven months (6.4 years) for the sixty-nine respondents to the question. Three CPAs chose not to respond to the question.

Most of the auditors who participated in the study had taken either one or two courses in statistics. Only four of the respondents indicated that they had not taken any statistic courses. The majority (i.e., 41) of the auditors had taken one course in statistics, while only 11 auditors responded that they had taken more than two statistic courses.

Table 3 (p. 176), reflects the number of years that the auditors have been associated with their present firm. Approximately, 74 percent of the auditors in this study have greater than three years experience with their present firm.

TABLE 3

Years Associated with Present Firm

Years	Number	Per Cent
3 or less	20	25.6
4 - 5	19	24.4
6 - 8	22	28.2
9 - 10	5	6.4
11 - 12	4	5.1
13 - 15	5	6.4
over 15	3	3.9
Total	78	100.0

Only seven of the respondents have been associated with their present firm for a period of a year or less. Since this study was geared to the local and regional public accounting firms, which rely heavily on nonstatistical sampling, these results suggest that on average, the auditors who participated in the experiment have had ample time to become acclimated to their firm's nonstatistical sampling approach.

For the seventy-eight auditors who participated in the study, the mean amount of public accounting experience was 8.1 years. Since in most local and regional public accounting firms the auditor can be expected to routinely render nonstatistical sample size decisions after four years of experience, the results in Table 4 (p. 178) are important. Approximately 89 percent of the respondents possessed the level of expertise at which they could be expected to make nonstatistical sample size decisions.

TABLE 4

Auditing Experience

Years Auditing Experience	Number	Per Cent
3 or less	9	11.5
4 - 6	31	39.7
7 - 8	11	14.1
9 - 10	6	7.7
11 - 15	13	16.7
16 - 20	5	6.4
over 20	3	3.9
Total	78	100.0

Question 10 asked the auditors to indicate the percentage of their total sampling applications that could be appropriately characterized as nonstatistical. To properly draw conclusions as to how auditors render nonstatistical sample size decisions, it is imperative that this sample of auditors routinely perform nonstatistical sampling. By selecting only local and regional public accounting firms for participation in this study, it was expected that the percentage performing nonstatistical sampling would be high. On average, 82.6 percent of the auditors' sampling applications could be characterized as nonstatistical. Forty-two (53.8 percent) practitioners indicated that 100 percent of their sampling was nonstatistical.

5.1.3 Attitudinal Characteristics

Question 11 asked the auditors to indicate if they had taken any continuing professional education (CPE) courses or seminars on the topic of audit sampling. Thirty-four (44.2 percent) auditors responded in the affirmative.

Question 12 asked the auditors to indicate their level of understanding of SAS NO. 39. Table 5 (p. 180), represents the breakdown of responses given by the auditors to this question. While a majority of the respondents indicated an adequate, working-knowledge of the Statement, twenty-

TABLE 5

Understanding of SAS No. 39

Level of Understanding	Number	Per Cent
Complete	3	3.8
Adequate (working-knowledge)	51	65.4
Insufficient working knowledge	22	28.2
No understanding	2	2.6
Total	78	100.0

four auditors (30.7 percent) were of the opinion that their understanding of SAS No. 39 was insufficient. When one considers that SAS no. 39 was initially issued as an exposure draft more than 3 years ago, the number of auditors who indicated a lack of understanding of the Statement is particularly surprising and disturbing.

SAS No. 39 provides standards that the auditor should consider when performing nonstatistical sampling in substantive tests of details. For example, paragraph 16 states specifically that the auditor should consider : (1) tolerable error, (2) risk of incorrect acceptance, and (3) characteristics of the population.[147] Thus, it is apparent that the Statement is designed to have a strong influence on auditors' nonstatistical sampling applications.

In response to question 13 (What effect has SAS No. 39 had on your nonstatistical sampling application?), thirty-two (41 percent) of the seventy-eight respondents indicated that the Statement has little or no effect on their nonstatistical sampling (see Table 6, p. 182). It would appear that a substantial portion of auditors are choosing to ignore the Statement.

---

[147] American Institute of Certified Public Accountants, "Audit Sampling," Statement on Auditing Standards, No. 39, (New York: AICPA, 1981), para. 16.

TABLE 6

Effect of SAS No. 39

Effect	Number	Per Cent
Strong	15	19.2
Moderate	31	39.7
Little or no	32	41.1
Total	78	100.0

Finally, question 14 asked the auditors to describe their general reaction to SAS No. 39. The respondents were asked to circle the letter(s) which correspond to the statement(s) describing their reaction. The following choices were offered:

A. The Statement provides needed guidance and direction.

B. The Statement is too technical for my nonstatistical sampling tasks.

C. The Statement unduly influences me to adopt a statistical sampling approach.

D. The Statement is a classic example of standards overload.

E. Other (explain)____

The results to question 14 are predictable, particularly when one considers the previous responses to questions 12 and 13. Less than half of the auditors (i.e., 35), expressed an opinion that SAS no. 39 provided needed guidance and direction. Table 7 (p. 184), reflects the results to question 14.

TABLE 7

Reaction to SAS No. 39

Reaction (Letter)	Number	Percent
A. (Provides Guidance)	35	44.9
B. (Too technical)	11	14.1
C. (Influences me to statistical)	6	7.7
D. (Standard overload)	12	15.4
E. (Other)	16	20.5

Note: since respondents could select more than one response, the percentages total more than 100.

The following is a sampling of the reactions that were expressed by the respondents who circled the letter 'E' (other):

...The Statement provides needed guidance and direction but is lacking in suggesting practical applications.

...(The Statement) is not specific enough.

...I believe there is still confusion within the profession as to when and where nonstatistical sampling is appropriate--i.e., the Statement needs additional interpretation.

...Too vague with weak conclusions.

...Statement is not appropriate for the majority of audit engagements a local CPA firm undertakes.

5.2 RESULTS OF STATISTICAL TESTS

5.2.1 Multiple Regression

To determine the factors that are used by the typical auditor in rendering nonstatistical sample size decisions, a group analysis of judgments made by seventy auditors was performed. From a total of fifty audit cases that were prepared for the study, each auditor was randomly assigned five cases for which sample size decisions were to be made. A random permutation of the cases insured that each case was analyzed once by seven different auditors.

To infer which factors are used by the typical auditor in making nonstatistical sample size decisions, the mean sample size recommendations for each of the fifty cases was regressed onto the independent factors. The resulting linear regression equation represents an estimate of the typical auditor's judgment strategy in nonstatistical sample size decisions.

A total of twelve independent variables or factors were used in the regression procedure. The following listing defines the predictor variables:

X(1)---Population variation (Sigma)

X(2)---Ultimate Risk (UR)

X(3)---Internal control risk (IC)

X(4)---Analytical review risk (AR)

X(5)---Risk of incorrect acceptance (TD)

X(6)---Tolerable error (TE)

X(7)---Population Size (N)

X(8)---Expectation of error (E)

X(9)---Risk of incorrect rejection (Alpha)

X(10)--Last year's sample size (Past)

X(11)--Account receivable dollar value (Asset)

X(12)--Total year's sales (Sales)

All but variables X(10) through X(12) are described in the authoritative literature that addresses the topic of audit sampling. The independent variable 'Past' (X(10)) was included in each case to determine whether auditors tend to anchor this year's sample size recommendation around last year's number. This would imply that auditors were ignoring pertinent information that is available in the form of the other given variables. The cases were constructed so that the correlation between the variable 'Past' and the cases' 'mean estimation' (optimal) sample size was a nonsignificant .091.

The variables 'Asset' and 'Sales' were included in the regression procedure when the pilot study's results revealed that such variables might act as explanatory factors. Again, the cases were prepared so that there were nonsignificant correlations (-.21 and -.03 respectively) between these factors and the criterion 'mean estimation' sample size.

Table 8 (p. 188) contains the results of the regression analysis for the group study. The upper segment of the printout consists of the analysis of variance table for testing the significance of the regression model.

TABLE 8

Regression Analysis of Group Study

DEP VARIABLE: SAMPLE SIZE

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	p(F)
MODEL	12	11035.001	919.583	7.068	0.0001
ERROR	37	4813.819	130.103		
TOTAL	49	15848.820			
			R-SQUARE	.70	

VARIABLE	DF	REGRESSION WEIGHT	t	p(t)
INTERCEPT	1	-28.26	-1.345	0.1869
SIGMA	1	-0.03	-0.658	0.5146
UR	1	39.68	0.458	0.6494
IC	1	17.80	1.505	0.1409
AR	1	15.89	1.474	0.1489
TD	1	15.19	0.327	0.7458
TE	1	0.00125	1.259	0.2160
N	1	0.00216	0.108	0.9145
E	1	-0.0064	-1.831	0.0752
ALPHA	1	113	1.844	0.0733
PAST	1	0.48	6.594	0.0001
ASSET	1	0.00011	2.366	0.0233
SALES	1	.0000084	1.426	0.1622

The significant F (7.07, p < .0001) indicates that the case factors account for a significant proportion of recommended sample size variance. The multiple R-squared of .70 (R = .83) indicates that the twelve independent variables account for 70 percent of the variance of recommended sample size.

Table 8 can also be used to assess which factors are used by auditors in making nonstatistical sample size decisions. The least squares regression weights for each factor are shown in the column labeled 'REGRESSION WEIGHTS'. These weights reflect the relative influence of the factors on the typical auditor's judgment policy. The 't' and 'p(t)' columns indicate the statistical significance of the regression weights for each independent variable.

Only variables X(10) (Past) and X(11) (Asset) have regression weights significantly different from zero. Of these two, the factor 'Past' is decidedly more significant with a p-value of 0.0001. Subject to the limitations of the sample, it would appear that the typical auditor is failing to consider most of the variables that are defined in SAS No. 39 and the related Audit Sampling Guide, and is focusing primarily on the sample drawn in the previous year. The 'Past' factor alone yields an R-square of .43.

Since auditors apparently use inappropriate information in making nonstatistical sample size judgments, and overlook relevant factors, it is crucial to ask whether the typical auditor is (nevertheless) making optimal or ideal sample size decisions relative to the 'mean estimation' statistical model. To evaluate this question, a t-test was performed to test the null hypothesis that the average difference between the cases' mean recommended sample sizes and the 'mean estimation' sample sizes is equal to zero. The average sample size difference per case (recommended - ideal) was 16.52. This mean yielded a t-value of 4.405 ($p <. 005$), indicating that the typical auditor is not making optimal or ideal sample size decisions relative to the 'mean estimation' statistical model.

The mean difference figure of 16.52, suggests that the typical auditor's sampling approach is inefficient. Subject to the limitations of the restrictive sample, the conclusion can be drawn that the typical auditor is over-auditing with respect to his sampling applications by approximately 56 percent.

Further insight into the typical auditor's sampling approach was gained by computing the correlation between the auditors' mean sample size recommendations and the 'mean estimation' sizes across the fifty cases. This correlation equaled 0.031 (n.s.); thus, we can conclude that there is no evidence of a relationship between the factors that the typical auditor relies on in making sampling decisions, and the factors of the 'mean estimation' statistical model.

Figure 7 (p. 192), represents a plot of the auditors' mean recommended sample sizes (symbol = B), and the 'mean estimation' sizes (symbol = A) across the fifty cases. On the average, the plot reveals that the typical auditor is grossly overauditing in his sampling judgments. In 80 percent of the cases, the auditors' mean recommendations were greater than the 'mean estimation' sample sizes. In two of these situations, cases 38 and 45, the auditors' recommendations exceeded the 'mean estimation' sizes by more than 500 percent. Thus, on average the typical auditor's nonstatistical sampling application though audit effective, is very inefficient.

In nine of the cases (18 percent); however, the auditors' mean recommendations were smaller than the 'mean estimation' sizes (e.g., cases 13, 16, 25). This suggests that although auditors are on average over-auditing, there are a fair number of instances where their nonstatistical sample size judgments are ineffective

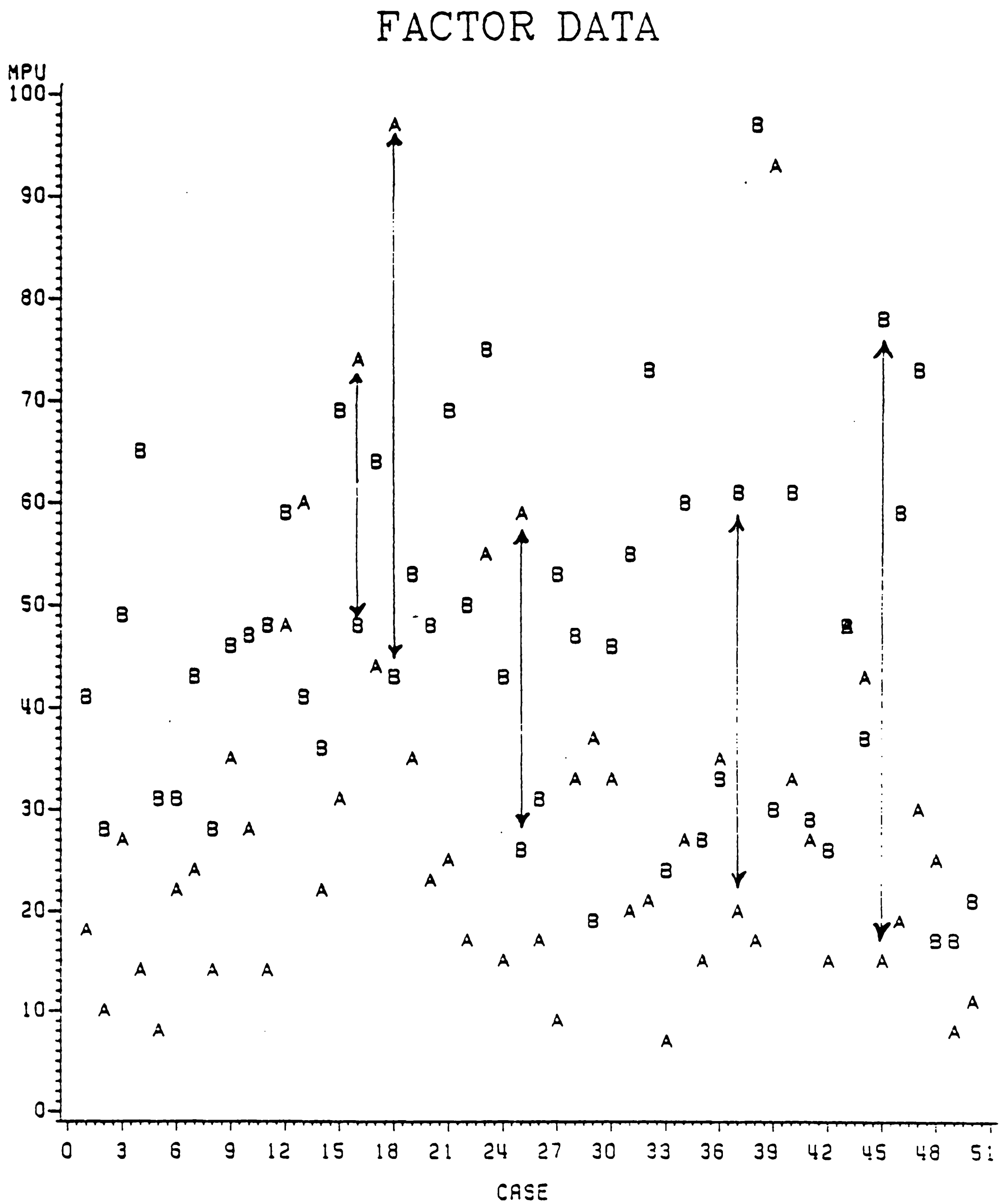


FIGURE 7

Factor Data

5.2.2 Analysis of Variance

The third major objective of this study was to determine whether significant differences exist between different auditors' nonstatistical sample size decisions. To evaluate this question, the fifty fictitious audit cases were all analyzed by each of eight auditors in a completely crossed factorial design.

Table 9 (p. 194), presents the results of this analysis. The significant 'Cases' main effect indicates that, as expected, different sample sizes were recommended for the fifty cases. Subject to the limitations of the sample, the highly significant effect for 'Auditors' indicates that auditors are making significantly different average sample size recommendations. Moreover, 44.6 percent of the total variance in sample size judgments is attributable to auditors.

TABLE 9

Anova Summary Table for Between Auditor Analysis

SOURCE OF VARIANCE	DF	SUM OF SQUARES	F	P
CASES	49	54678.89	3.65	.0001
AUDITORS	7	128650.35	60.12	.0001
ERROR	343	288188.39		

A further indication of how consistent the auditors' average sample size recommendations are can be obtained by computing the intraclass correlation. The intraclass correlation evaluates the agreement or consistency in patterns of recommended sample sizes between the auditors' judgments across the fifty cases. The correlation is given by the following formula:

$$r_{cc} = \frac{(MS)r - (MS)e}{(MS)r + (k-1)(MS)e}$$

where (MS)r = mean square or variance between cases.

(MS)e = mean square for residuals, or error.

k = number of auditors.

For the data of Table 4, $r_{cc} = .249$.

This value can be interpreted as the expected correlation between sample size vectors for any pair of auditors randomly chosen from the population of auditors from which this study's sample was selected. This .25 inter-judge reliability is inadequate.

The extent of variability in auditors' mean sample size recommendations can be observed in Table 10 (p. 197). Table 10 reflects the results of Duncan's Multiple Range Test comparing the means of the auditors' recommended sample sizes. Duncan's test controls the type I comparisonwise error rate (not the experimentwise error rate) at .05. In Table 10, mean sample sizes with the same letter are not significantly different. The results of the Duncan procedure confirm that auditors' mean sample size decisions are extremely variable.

In evaluating the results of the foregoing statistical analysis, it is important to keep in mind that the respondents in the study were not student surrogates, or junior auditors who lacked the professional expertise to render nonstatistical sample size decisions. The eight auditors who participated in this phase of the experiment, averaged more than seven years auditing experience, with only one auditor having less than five years experience (four years). Furthermore, the respondents indicated that on average, 97 percent of their sampling applications could be appropriately classified as nonstatistical. Viewed in this framework, the study's findings are a matter of concern for the public accounting profession.

TABLE 10

Duncan's Multiple Range Test

GROUPING		MEAN	N	AUDITOR
	A	90.38	50	4
	B	63.64	50	2
C	B	57.36	50	6
C		52.58	50	8
	D	45.00	50	5
	D	40.84	50	3
	E	32.56	50	1
	E	32.40	50	7

Having gathered empirical evidence that significant differences exist between auditors' mean sample size decisions attention is now focused on identifying those variables that best explain such differences. The first step in this process involved a separate regression of each auditor's fifty recommended sample sizes onto the independent variables. This yielded a significant multiple R for each of the eight respondents. The results are presented in Table 11 (p. 199).

To identify the factors that best accounted for each auditor's mean sample size decisions, a backward stepwise regression was performed for each auditor's recommendations. A listing of those variables that received significant regression weights in the auditors' regression models at the 0.01 level is given in Table 12 (p. 200). The factors 'TE', 'IC', 'AR', and 'Asset' were significant most frequently.

TABLE 11

Auditor R-Square

Auditor	R-Square	F	PROB>F
1	.792	11.73	0.0001
2	.769	10.27	0.0001
3	.659	5.95	0.0001
4	.755	9.48	0.0001
5	.620	5.04	0.0001
6	.867	20.12	0.0001
7	.727	8.23	0.0001
8	.525	3.40	0.0020

TABLE 12

Explanatory Factors

Auditor	Variable	F	Prob>F
1	TE	86.47	0.0001
	N	14.33	0.0005
2	TD	51.94	0.0001
	TE	27.45	0.0001
	E	13.79	0.0006
3	IC	27.11	0.0001
	N	8.89	0.0047
4	IC	43.38	0.0001
	E	25.89	0.0001
	ASSET	18.30	0.0001
5	ASSET	43.16	0.0001
6	IC	125.27	0.0001
	AR	41.12	0.0001
	TE	13.30	0.0007
7	TE	39.19	0.0001
	AR	17.46	0.0001
	ASSET	15.33	0.0003
8	TE	13.07	0.0008
	ASSET	11.84	0.0013
	AR	10.92	0.0019

where: Asset = total accounts receivable value
TE = tolerable error
TD = risk of incorrect acceptance
N = population size
E = expectation of error
IC = internal control
AR = analytical review

Table 13 (p. 202), depicts the correlations and p-values between each of the four factors and the criterion variable 'Size' for each of the eight respondents. A review of Table 13 underscores the diversity with which the respondents relied upon these independent variables to render their sample size judgments. For example, there was a positive significant correlation between the variables 'TE' and 'Size' for auditor 1, but, the same two variables had a significant negative correlation for auditor 7. Similarly, there was a positive significant correlation between the variables 'Asset' and 'Size' for auditors 1, 4, and 5, but this correlation was very low and nonsignificant for auditors 2 and 8, and was significantly negative for auditor 7.

TABLE 13

Correlations between Four Factors and Judgments for Eight Auditors

Auditor	Factors			
	TE	IC	AR	Asset
1	.81	.14	.22	.73
2	-.20	.35	.47	-.08
3	.29	.58	.40	.28
4	-.14	.68	.28	.52
5	.70	.32	.15	.70
6	.02	.79	.59	.07
7	-.47	.28	.41	-.28
8	-.09	.33	.37	.08

Note: Correlations are computed across fifty cases. Critical values for 'r' are:

$p < .01 = .36$

$p < .05 = .28$

$p < .10 = .24$

$p < .20 = .18$

$p < .50 = .10$

5.3 INDICES OF JUDGMENT

5.3.1 Judgmental Accuracy

As noted in Chapter 3, the accuracy of auditors' judgments has received little attention in the decision making literature. Since no definitive criteria presently exist for defining the optimal internal control system or what constitutes the threshold of materiality, the accuracy of subjects' judgments has not been an issue in studies that have focused on internal control evaluation or materiality judgments. Thus, the judgment performance measure to attract the greatest focus in these areas has been judgment agreement or consensus.

Judgmental accuracy has been found to be wanting across a multitude of diverse judgment tasks and experimental conditions.[148] In those studies which have assessed judgmental accuracy in the clinical judgment literature, Ashton reports that accuracy is typically quite low and is largely unaffected by the amount of information that is presented to the individual or the amount of professional experience or training that the rater possesses.[149] Instead, the elements of the judgment task appear to affect the level of accuracy

---

[148] Ashton, Robert H., Human Information Processing in Accounting, Studies in Accounting Research No. 17 (Sarasota, Florida: American Accounting Association, 1982), p. 30.

[149] Ibid, p. 30.

that is attained. For example, accuracy is higher in those studies where the independent variables are linearly related to the criterion than when they are nonlinearly related.[150]

In the present study, since an optimal or ideal sample size can be determined for each case, research question 2 (How accurate are the nonstatistical sample size recommendations of the auditor when compared to the 'mean estimation' statistical sample size?) can best be evaluated by aggregating the auditors' sample size predictions on the fifty cases. With each case being analyzed once by fifteen different auditors, judgmental accuracy is simply the average correlation between the auditors' recommended nonstatistical sample sizes and the 'mean estimation' statistical sample sizes over all of the cases. Thus, the accuracy correlation will be based on 750 (15 per case X 50 cases) pairs of observations.

Earlier in this chapter we found that the typical auditor was ignoring most of the pertinent information that was supplied to him in the form of the independent variables. Specifically, the typical auditor was relying largely on the variables, 'Past' and 'Asset' in rendering nonstatistical sample size decisions; the remaining variables were virtually excluded from the decision process. Such inappropriate

---

[150] Ibid, p. 30.

judgment behavior, as we shall see, had a detrimental effect on judgmental accuracy.

In this study's other phase that involved eight auditors each analyzing all fifty cases in a factorial design, we concluded that auditors are making significantly different average sample size decisions averaging across cases. Furthermore, the empirical evidence indicates that approximately 44.6 percent of total variance in the cases' sample size judgments, can be explained by judgmental differences among auditors. Thus, from both phases of this study we acquire an appreciation of the inconsistency that exists among the auditors' sample size decisions.

Consistent with these earlier findings, the accuracy of auditors' judgments was assessed to be extremely low, with a correlation of 0.01 (p-value >.78). Subject to the restrictive nature of the sample, these results should be interpreted with caution.

The average recommended nonstatistical sample size was 49.14, with a standard deviation of 29.57. In comparison, the average 'mean estimation' size was 29.54, with a standard deviation of 19.83. Thus on average, the seventy-eight auditors who participated in this study, are over-auditing in their sample size decisions, by approximately 66 percent. The 'mean estimation' sample sizes ranged from 7 to 97. The auditors' recommended sizes ranged from 0 to 200.

The amount of auditing experience possessed by the respondents had little effect on accuracy. For the forty auditors who each had less than or equal the median experience level of six years, the accuracy index was 0.02 (p-value >.63). For the thirty-eight auditors whose level of professional expertise exceeded the median amount, accuracy was evaluated to be 0.004 (p-value >.94).

It was expected that auditors who indicated on the questionnaire that they had taken courses or seminars which focused on audit sampling, would have judgmental accuracy exceeding the overall composite index of 0.01. While this in fact was the case, the difference was not significant, and the accuracy of these auditors' sample size decisions remained extremely low ($r = .02$, $p > .62$). Thus there was no evidence that these auditors had benefited significantly from their audit sampling instruction. The accuracy index for the forty-three auditors who had not taken any CPE courses or seminars on sampling, was 0.008 (p-value >.87).

It was hypothesized that auditors who indicated at least an adequate understanding of SAS No. 39 and who felt that the Statement had at least a moderate effect on their nonstatistical sample size decisions, would attain a greater level of judgmental accuracy than their colleagues who did not respond positively to these questions. To evaluate this hypothesis, each auditor's root mean square error in estimating required sample sizes was correlated with his responses to questions 12 and 13. This approach yielded an N of 78.

The correlations between the auditors' root mean square errors and the responses to questions 12 and 13 were 0.04 (p-value >.71) and -0.05 (p-value >.67). Thus, we fail to reject the null hypothesis and conclude that auditors who indicated at least an adequate understanding of SAS No.39 and who responded that the Statement had at least a moderate effect on their sampling behavior, do not achieve significantly greater judgmental accuracy than those auditors who acknowledged less than an adequate understanding of SAS No. 39 and who felt that the Statement has little or no effect on their nonstatistical sampling applications. The evidence suggests that, on average, auditors' assessment of their knowledge of SAS no. 39 is over-stated and unjustified.

From a number of perspectives, the 'accuracy' of the auditors' nonstatistical sample size decisions has been found to be extremely low and nonsignificant. A possible explanation for these findings might be the difficulty that auditors experienced in trying to consider (implicitly) all of the pertinent factors that SAS No. 39 enumerates. As noted previously in this section, judgmental accuracy has been shown to be unaffected by the amount of professional experience or training that the individual has acquired. Rather, it is the elements of the judgment task, and how they are interrelated with each other that appear to have the most significant impact on accuracy.

5.3.2 Judgment Agreement (Consensus)

In the audit decision-making literature, consensus has been the judgment performance measure that has received the greatest research attention. Ashton defines consensus as agreement across individuals using the same data at one point in time.[151]

In the behavioral research that was reviewed in Chapter 3, the consensus among individuals' predictions was found to be strong. In those studies that dealt with auditors' evaluation of internal control, the level of consensus typically ranged from .60 to .70. Similar results on the consensus dimension were found in the bankruptcy prediction research. For example, Kida found a consensus of .755 among auditors' judgments.[152]

---

[151] Ibid, p. 45.

[152] Kida, Thomas, "An Investigation into Auditors' Continuity and Related Qualification Judgments," Journal of Accounting Research (Autumn 1980), p. 519.

Much evidence exists that indicates high levels of consensus are linked to the number of factors that comprise the judgment task. For example, in the internal control evaluation studies, generally six dichotomously-scaled factors were presented to the auditors. Similarly, there were usually five or six financial ratios that served as factors in the research that focused on the prediction of firm failure.

In the present study, research question 3 (i.e., What is the level of agreement across auditors' nonstatistical sample size judgments?) was evaluated by computing the intraclass correlation reported above. This consensus estimate was .25 ($p > .08$). This is marginally nonsignificant at the .05 level. When analyzed in the context of the consensus findings which have been reported in the areas of internal control evaluation and bankruptcy prediction, these results are not particularly impressive. Again, interpretation of these results must be considered in light of the limitations of the sample.

5.3.3 Agreement in Principle

Correlational analysis was used to evaluate research question 4 (i.e., How similar are different auditors' judgment policies for making nonstatistical sample size decisions?). For each of the eight auditors who evaluated all fifty cases, their predicted sample size judgments based on their regression equation were determined. Agreement in principle was measured by correlating the predicted responses of all subject pairs across each of the fifty cases and taking an average.

With a correlation of .425 (p-value <.002), we reject the null hypothesis that there is no association between the eight auditors' sample size judgment policies.

5.3.4 Auditor Self-Insight

The fifth and final research question (How much self-insight do auditors exhibit in their judgment policies?) was evaluated with a Spearman rank correlation coefficient. On average, the association between the ranks of the standardized regression coefficients for each of the eight auditors who evaluated all fifty cases and the ranks of these auditors' subjectively assigned beta weights was 0.49 (p-value >.10). Thus, at the 0.10 level we are unable to reject the null hypothesis of no association between the two sets of rankings. We conclude that on average, the auditors did not exhibit significant insight into their judgment policies.

These results on the insight dimension are consistent with the findings reported by Ashton.[153] That is, the degree

[153] Ashton, R., Human Information Processing in Accounting, Studies in Accounting Research No. 17, p. 40.

of association between individuals' subjective weights and statistically derived weights is generally quite low. In this particular study the auditors generally were of the opinion that they utilized more of the independent variables than was indicated by their linear multiple regression equations. Furthermore, there was a definite tendency for the auditors to overweight variables that were relatively unimportant in their regression equations.

This evidence suggests that if the public accounting profession is to revamp the nature of nonstatistical sampling to make it more audit effective and efficient, a well-organized, concerted, educational effort is required. Presently, auditors display no significant insight into how they render their nonstatistical sample size decisions. To quote a phrase which was heard often by this researcher during the administration of the experiment, auditors tend to "fly by the seat of their pants" when they render nonstatistical sample size judgments.

5.4 CHAPTER SUMMARY

The empirical evidence that has been gathered in this study represents a potential matter of concern to the public accounting profession. The typical auditor, in his nonstatistical sample size decisions, ignores the factors that SAS

No. 39 enumerates. Last year's sample size ('Past') is clearly the most explanatory variable in decisions of this type. In electing to ignore SAS No. 39, the typical auditor is not making ideal or optimal sample size judgments.

Eight auditors who each analyzed all fifty cases provided powerful evidence that auditors are making statistically different average sample size recommendations averaging across the cases. In a completely crossed factorial design, approximately 44.6 percent of the variance in sample size judgments was attributable to the 'auditor' effect.

Since auditors apparently do not use the guidance provided by SAS no. 39, it was not surprising that their sample size judgments were not significantly correlated with the 'mean estimation' statistical sample sizes. The judgmental accuracy of the auditors was only 0.01 (p-value >.78). What was surprising was that auditors who indicated at least an adequate understanding of SAS No. 39, and who felt that the Statement has at least a moderate effect on their nonstatistical sample size judgments did <u>not</u> attain significantly greater judgmental accuracy than that of their colleagues who professed a less than adequate understanding of SAS No. 39, and who were of the opinion that the Statement has little or no effect on their nonstatistical sampling behavior.

In the audit decision-making literature, the consensus of individuals' judgments has generally been found to be statistically strong (e.g., .60 to .70). Such was not the finding in this particular study. The intraclass correlation revealed that any pair of auditors chosen from the population of auditors from which this study's sample was selected, would be expected to have their nonstatistical sample size judgments correlate only .25 across the fifty cases. Thus, auditors are apparently using a varied assortment of techniques and procedures to render highly inaccurate and inefficient nonstatistical sample size decisions.

Finally, the evidence indicates that auditors possess no significant self-insight into how they make their sample size judgments. That is, auditors do not display a knowledge of the factors that are statistically most influential to their judgmental processes.

Chapter VI

SUMMARY AND CONCLUSIONS

SAS No. 39 defines audit sampling as the application of an audit procedure to less than 100 percent of the items within an account balance or class of transactions for the purpose of evaluating some characteristic of the balance or class.[154] The rationale that supports sampling or testing as a well established audit technique is that some samples are sufficiently representative to warrant valid and reliable conclusions about the entire population of transactions or balances under investigation.[155] Today, auditors place increasingly greater reliance upon sampling procedures as they seek sufficient, competent evidential matter within business entities that continue to grow in size and complexity.

---

[154] Statement on Auditing Standards No. 39, Audit Sampling, (New York: AICPA, 1981), paragraph 1.

[155] Ricchiute, David N., Auditing: Concepts and Standards (South-Western Publishing Co., 1982), p. 296.

6.1 SUMMARY OF THE PROBLEM

Audit sampling can be either statistical or nonstatistical according to SAS No. 39. The primary difference between the two sampling methods is that statistical sampling measures and controls the risk of material sampling error. SAS No. 39 least effects those large public accounting firms that continue to use statistical sampling procedures. However, the Statement continues to be a source of controversy for those practitioners who employ nonstatistical sampling.

SAS No. 39 discusses audit sampling in the context of two general approaches, namely: (1) compliance tests of internal accounting controls, and (2) substantive tests of details. The focus throughout this experimental study has been on auditors' nonstatistical sample size judgments in substantive tests of details. This emphasis on substantive tests of details corresponds to the substantive audit approach that is typically used by local and small regional firms in their audits of small businesses.

Since its issuance in June 1981, SAS No. 39 has aroused the concerns of those auditors who use nonstatistical sampling as they experience difficulties in designing, selecting, and evaluating a nonstatistical sampling procedure in accordance with the Statement. For example, SAS No. 39 stipulates that auditors should consider the following factors when planning a particular sample for a substantive test of details:

(1) Preliminary estimates of materiality levels.

(2) The auditor's allowable risk of incorrect acceptance.

(3) Characteristics of the population, that is, the items comprising the account balance or class of transactions of interest.[156]

All of the above factors are explicitly considered in statistical sampling applications by their inclusion in mathematical models. However, auditors who use nonstatistical sampling must draw upon their judgment and professional experience to subjectively analyze these factors and their effect on sample size.

In the course of an audit engagement, auditors must make numerous judgments and express an opinion on the basis of the information and test results that are never perfectly diagnostic with respect to the underlying economic state of the client. The auditor's decision as to sample size in substantive tests of details is but one example of such a judgment. The ability of auditors to subjectively formulate appropriate judgments is crucial, since they may be held

---

[156] Statement on Auditing Standards No. 39, Audit Sampling, paragraph 16.

liable at common law or under the federal securities laws should the audited financial statements prove to be materially in error. Thus, this threat of litigation accentuates the importance of improving the effectiveness of audits. Additionally, subjective decisions on sample size impact the efficiency of audits. As competition among firms for clients continues to grow, there is an increasing emphasis on improving the efficiency of audits. The relevance and significance of the audit sampling issue, specifically auditors' determination of nonstatistical sample size decisions is underscored when one considers the importance of audit effectiveness and audit efficiency.

6.2 EMPIRICAL FINDINGS

The experimental study involved both a group and individual analyses. Seventy practitioners particpated in the group segment of the study that was designed primarily to ascertain what factors are used by the typical auditor in making nonstatistical sample size judgments. In order to provide evidence in support of this objective, the mean sample size recommendations for each of fifty fictitious cases were regressed onto the independent factors.

Both the full model regression solution and forward and backward stepwise analyses indicated that the variable

'Past' (i.e., last year's sample size), was by far the most explanatory factor in the auditor's judgment. That is, when the forward inclusion technique was run, the variable 'Past' was first to enter the model accounting for 62 percent of total explainable auditor judgment variance. This finding suggests that auditors tend to ignore the factors that SAS No. 39 enumerates. Simply put, the nonstatistical sample sizes that the typical auditor selected this period were influenced almost exclusively by the sizes chosen last year, regardless of the occurrence of significant changes in factors such as risk of incorrect acceptance, tolerable error, etc..

Having gathered evidence as to the factors that are used by the typical auditor in making nonstatistical sample size judgments, the following related question was posed: Is the typical auditor rendering optimal or ideal nonstatistical sample size judgments on the cases relative to the 'mean estimation' statistical model? To evaluate this question, the methodology involved a t-test to test the null hypothesis that the average difference between the cases' mean recommended nonstatistical sample sizes and the 'mean estimation' sizes is equal to zero. The null hypothesis was rejected at the 0.0001 level of significance. Furthermore, the mean difference figure of 16.52 provides evidence that the typical auditor's nonstatistical sample size judgment is inefficient.

The second major phase of this study involved a sample of eight auditors each of whom analyzed all fifty cases. This individual analysis was designed to determine whether significant differences exist between auditors' nonstatistical sample size judgments.

An analysis of variance (ANOVA) design provided evidence that auditors are making statistically different mean sample size recommendations averaging across cases. A backward elimination routine was run in an attempt to determine the factors that might best account for the significant differences between auditors' mean nonstatistical sample size judgments. The backward analysis identified the following four factors: (1) internal control risk, (2) analytical review risk, (3) tolerable error, and (4) the dollar value of the accounts receivable. Thus, it was the auditors' diverse use of these factors that best explains the significant differences in their nonstatistical sample size recommendations.

This study also assessed auditor judgmental accuracy, consensus, and self-insight. Across all of these indices of judgment the empirical results are very low.

Judgmental accuracy was evaluated by taking the average of each case's fifteen nonstatistical sample size judgments and correlating these values with the 'mean estimation' sizes. With a correlation of 0.01 (p-value >.78), we failed to reject the null hypothesis that auditor judgmental accuracy was not significantly different from zero. This finding came as no great surprise, particulary after evidence had been gathered that indicates auditors fail to consider the factors that SAS No. 39 enumerates.

The decision-making literature is replete with evidence that indicates that accuracy is a function of the number of factors that the individual must manipulate in the judgment task. Since numerous factors must be considered by auditors in their nonstatistical sample size judgments, herein is one possible reason for the nonsignificant finding on the accuracy dimension.

The consensus dimension was evaluated by the intraclass correlation. The intraclass correlation evaluates the agreement or consistency in patterns of recommended sample sizes between the auditors' judgments across the fifty cases. For the data in the study, the intraclass correlation was .249. This value can be interpreted as the expected correlation between sample size vectors for any pair of auditors randomly chosen from the population of auditors from which this study's sample was selected. This .249 interjudge reliability is inadequate.

The final index of judgment, which was evaluated in this study, was auditor self-insight. Specifically, the Spearman rank correlation coefficient was used to assess the association between the ranks of the standardized regression coefficients for each of the eight auditors who evaluated all fifty cases, and the ranks of these auditors' subjectively assigned beta weights. At the 0.10 level of significance, we were unable to reject the null hypothesis of no association between the two sets of rankings and accordingly concluded that on average, auditors did not exhibit significant insight into their nonstatistical sample size judgment process. That is, auditors did not display significant knowledge of those factors that influence and control their judgment process relative to nonstatistical sample size judgments. To quote a phrase that was heard often by this researcher during the administration of the experiment, auditors tend to "fly by the seat of their pants" when they render nonstatistical sample size judgments.

In evaluation of the foregoing empirical findings, it important to remember that the respondents in this study were not student surrogates or junior auditors who lacked the professional expertise to render nonstatistical sample size judgments. On average, the seventy-eight practitioners who participated in the study possessed slightly better than eight years auditing experience. Furthermore, these auditors indicated that on average, 83 percent of their testing applications could be characterized as nonstatistical.

6.3 CONCLUSIONS AND RECOMMENDATIONS

A basic proposition of this experimental study is that auditors have a professional responsibility to understand and follow the guidance provided by SAS No. 39 in their nonstatistical sample size judgments. However, the empirical evidence developed in this study supports the conclusion that practitioners have failed to comprehend the Statement's standards. Alternatively, this study has provided ample evidence that indicates auditors have chosen to ignore SAS No. 39.

In order for SAS No. 39 to eventually realize the impact on audit sampling for which it was intended, research that will enhance a conceptual understanding of audit sampling needs to be undertaken and promoted. Auditors need not only the procedural skill to effectively and efficiently perform audit sampling, they also require a conceptual understanding of sampling concepts. Only when auditors comprehend the theory upon which sampling is based, will they be able to consistently render optimal nonstatistical sample size judgments.

The auditors who participated in this study provided no empirical evidence that they understood some of the basic concepts associated with audit sampling. For example, the respondents demonstrated no significant appreciation for relating their risk of incorrect acceptance (i.e., beta risk) to reliance on internal accounting controls and analytical review procedures, including other substantive tests. Apparently, the typical auditor has failed to understand the model (i.e., TD = UR/(IC X AR)), for relating the risk components of an audit that has been provided in Appendix F of the Audit Sampling Guide. Although the auditors showed some utilization of the factors that pertained to the risk levels of internal accounting controls and analytical review procedures in their judgment process, this study's findings indicate that they failed to strongly consider their risk of incorrect acceptance in making nonstatistical sample size judgments. Furthermore, the respondents showed no understanding of the inverse relationship that exists between their risk of incorrect acceptance and their recommended nonstatistical sample sizes.

The results of this experimental study can be used to cite further examples of the auditors' failure to understand basic auditing concepts. In ten of the fifty cases, which were prepared for this study, all of the factors except two were kept constant. The two factors that varied throughout these cases were: (1) the auditor's risk of incorrect rejection (i.e., alpha risk), and (2) the variation in the population. In the majority of instances, the auditors rendered identical nonstatistical sample size judgments over all of these ten cases. Such findings provide evidence that auditors fail to discern both the direct effect that population variation has on sample size, and the inverse relationship between the auditor's risk of incorrect rejection and sample size.

The empirical evidence indicates that auditors fail to understand many of the basic relationships that exist between selected factors and their recommended sample size. For example, larger measures of tolerable error are known to lead to smaller sample sizes, while larger sample sizes are the result of smaller estimates of tolerable error. The results from this study reveal that the typical auditor did not discern this inverse relationship. Conversely, the factor, expectation of error, is known to have a direct effect on the auditor's recommended sample size. That is, the auditor's expectation of smaller errors leads to smaller sample sizes, whereas the auditor's anticipation of larger errors is a condition that directly effects larger recommended sample sizes. Such a direct relationship was not understood by the auditors in this experiment. To the contrary, the results of this study indicate that the respondents perceived an inverse relationship between expectation of error and sample size. The failure of practitioners to understand such fundamental relationships is indicative of their inability to comprehend the concepts related to audit sampling.

While research that will enhance auditors' conceptual understanding of audit sampling is urgently needed, the benefits of such efforts cannot reasonably be expected to be realized in the near future. Accordingly, some steps must be taken that will immediately affect auditors' nonstatistical sampling applications.

Throughout the administration of this experiment, many auditors expressed their desire for a type of 'sampling worksheet' that would enable them to document their compliance with SAS No. 39. Moreover, these auditors were of the opinion that such a worksheet would provide them with a procedural outline of SAS No. 39 that explicitly considers the factors that the Statement enumerates. As noted earlier in this study, a form of sampling worksheet is presently in use by the international public accounting firm, which participated in this study's pilot test. Furthermore, at least three of the larger regional firms that took part in this experiment, are in the process of instituting a worksheet approach to their nonstatistical sampling applications.

Proponents of an illustrative worksheet maintain that it will greatly reduce the diversity that is thought to exist among practitioners' nonstatistical sample size judgments. The extreme variability, which was observed among auditors' nonstatistical sample size judgments in this study, exemplifies the urgent need for an immediate remedy to the problem. Until auditors acquire a theoretical understanding of sampling concepts, a worksheet approach might tend to enhance uniformity among their nonstatistical sample size judgments.

6.4 LIMITATIONS OF THE STUDY

All of the seventy-eight auditors who participated in this study were employed by public accounting firms in the state of Virginia. The respondents were not randomly selected from the population of auditors, but were chosen based on the researcher's knowledge of their desire to participate. Although the sample was not randomly selected, it is representative of the population of auditors who primarily apply nonstatistical sampling techniques.

While we have evidence that the sample of auditors are making inefficient nonstatistical sample size judgments, the limitations of our sample could limit the generalization of the results to the population of auditors as a whole. However, the empirical evidence that has been gathered may be an indication that auditors who primarily render nonstatistical sample size judgments, are experiencing difficulties in their design, selection, and evaluation of a nonstatistical sample in accordance with SAS no. 39.

6.5 SUGGESTIONS FOR FURTHER RESEARCH IN AUDIT SAMPLING

The need for research that will enhance auditors' understanding of audit sampling concepts has been outlined in the previous section. In addition, research needs to be commenced that will provide a basis for understanding the reasoning behind auditors' sample size judgments.

Although the auditors in this study were not required to subjectively quantify the risks associated with internal accounting controls and analytical review procedures, since all of the factor levels had been pre-established by the author, such decisions are generally important inputs to the auditor's overall sample size judgment. Thus, research is needed that will identify the circumstances or conditions that would cause auditors to, for example, subjectively place moderate reliance on analytical review procedures and other substantive tests. Similarly, we need a better understanding of the conditions that influence the auditor's quantification of internal accounting control effectiveness. In short, studies are needed that will enhance our understanding of the conditions or circumstances that impact the levels of auditor confidence.

To date, a fair amount of research has been conducted that focuses on the extent to which auditors use heuristics in their professional judgment. Although the results of such research generally indicate the use of heuristics by auditors, the evidence is far from conclusive. The need for further research to examine auditors' use of heuristics in decision-making is underscored when one considers that the typical auditor in this study relied most on last year's sample size to render nonstatistical sample size judgments.

Finally, three years after its issuance, SAS No. 39 continues to be an object of heated debate and a source of controversy within the public accounting profession. Many of the practitioners who participated in this study displayed a contemptible disregard for the guidance provided by SAS No. 39. For example, an alarming number of respondents professed an inadequate knowledge of the Statement and an unwillingness to abide by its provisions. Accordingly, research is needed that will provide the profession with a theory of auditor attitude and behavior relative to official pronoucements, so that future Statements on Auditing Standards may be prepared in a manner that will enhance widespread acceptance and compliance.

BIBLIOGRAPHY

American Institute of Certified Public Accountants, _Auditing Standards and Procedures, Statement on Auditing Procedures No. 33_. New York: AICPA, 1963.

American Institute of Certified Public Accountants, "Generally Accepted Auditing Standards-Their Significance and Scope." _Special Report of Committee on Auditing Procedures_. New York: AICPA, 1954.

American Institute of Certified Public Accountants, "Statistical Sampling and the Independent Auditor." _Special Report by the Committee on Statistical Sampling_. New York: AICPA, 1962.

American Institute of Certified Public Accountants, Statistical Sampling Subcommittee. _Audit and Accounting Guide: Audit Sampling_. New York: AICPA , 1983.

Auditing Standards Board. "Audit Sampling". _Statement on Auditing Standards No. 39_. New York: AICPA, 1981.

Auditing Standards Board. "Evidential Matter." _Statement on Auditing Standards No. 31_. New York: AICPA, 1980.

Auditing Standards Board. "Planning and Supervision." _Statement on Auditing Standards No. 22_. New York: AICPA, 1978.

Auditing Standards Board. "Working Papers." _Statement on Auditing Standards No. 41_. New York: AICPA, 1982.

Arkin, H. _Handbook of Sampling for Auditing and Accounting_. New York: McGraw-Hill, 1974.

Ashton, R. H. "An Experimental Study of Internal Control Judgments." _Journal of Accounting Research_ (Spring 1974): 143-157.

Ashton, R. H. "Cue Utilization and Expert Judgments: A Comparison of Independent Auditors with Other Judges." _Journal of Applied Psychology_ (August 1974): 437-44.

Ashton, R. H. "The Predictive-Ability Criterion and User Prediction Models." _The Accounting Review_ (October 1974): 719-32.

Ashton, R. H. Human Information Processing in Accounting, Studies in Accounting Research No. 17 (Sarasota, Florida: American Accounting Association, 1982).

Ashton, R. H. "Some Implications of Parameter Sensitivity Research for Judgment Modeling in Accounting." The Accounting Review (January, 1979): 170-79.

Ashton, R. H. and P. R. Brown. "Descriptive Modeling of Auditors' Internal Control Judgments: Replication and Extension." Journal of Accounting Research (Spring 1980): 269-77.

Ashton, R. H. and S. Kramer. "Students as Surrogates in Behavioral Research: Some Evidence." Journal of Accounting Research (Spring 1980): 1-15.

Bailey, Andrew D. Jr. Statistical Auditing: Review, Concepts and Problems. (New York: Harcourt Brace Jovanovich, Inc., 1981).

Bailey, Larry P. "Impact of SAS-39 on Nonstatistical Sampling." The CPA Journal (June 1982): 38-47.

Boatsman, J. R. and J. C. Robertson. "Policy-Capturing on Selected Materiality Judgments." The Accounting Review (April 1974): 342-52.

Brown, Gene R. "Changing Audit Objectives and Techniques." The Accounting Review (October 1962): 696-703.

Biddle, G. C. and E. J. Joyce. "The Role of Sample Size in Probabilistic Inference in Auditing." Unpublished Manuscript, University of Chicago, 1979.

Casey, C. J. "The Usefulness of Accounting Ratios for Subjects' Predictions of Corporate Failure: Replication and Extension" Journal of Accounting Research (Autumn 1980): 603-13.

Carnegie Steele, F. R. "The Development of Systems of Control". The Journal Of Accountancy (1913): 282.

Chernoff, H. "The Use of Faces to Represent Points in N-Dimensional Space Graphically". Technical Report No. 71, Department of Statistics, Stanford University, December 1971.

Cochrane, George. "The Auditor's Report: Its Evolution in the USA". The Accountant (November, 1950): 448-460.

Committee on Auditing Procedure, American Institute of Certified Public Accountants. "Codification of Auditing Standards and Procedures." Statement on Auditing Standards No. 1. New York: AICPA, 1972.

Crosby, M. A. "Implications of Prior Probability Elicitation on Auditor Sample Size Decisions". Journal of Accounting Research (Autumn 1980): 585-93.

Danos, P., and E. A. Imhoff. "Auditor Review of Financial Forecasts: An Analysis of Factors Affecting Reasonableness Judgments". The Accounting Review (January 1982): 39-69.

Dawes, R. M., and B. Corrigan. "Linear Models in Decision Making". Psychological Bulletin 81 (1974): 85-106.

Dawes, R. M. and R. D. Diller. "The Prediction of Bootstrapping and a Method of Amalgamation". Oregon Research Institute Research Bulletin 10 (1970): 6-18.

Driver, M. J. and T. J. Mock. "Human Information Processing, Decision Style Theory, and Accounting Information Systems". The Accounting Review (July 1975): 490-508.

Edgell, S. E. "Configural Information Processing in Two-Cue Probability Learning". Organizational Behavior and Human Performance, (1978) Vol. 22: 404-416.

Goldberg, L. R. "Simple Models or Simple Processes"? American Psychologist (1968) Vol. 23: 483-496.

Goldberg, L. R. "Man Versus Model of Man: A Rationale, Plus Some Evidence, for a Method of Improving on Clinical Inferences". Psychological Bulletin (1970) Vol. 73: 422-432.

Hamilton, R. E. and W. F. Wright. "The Evaluation of Internal Controls over Payroll". unpublished manuscript, Stanford University, 1977.

Hammond, K. R., ed.. The Psychology of Egon Brunswik (New York: Holt, Rinehart, and Winston, 1966).

Hammond, K. R. and D. A. Summers. "Cognitive Dependence on Linear and Nonlinear Cues". Psychological Review (May 1965): 215-24.

Hays, W. L. Statistics (New York: Holt, Rinehart and Winston, 1973).

Hill, Henry, Joseph Roth and Herbert Arkin. Sampling in Auditing. (New York: Ronald Press, Inc., 1962).

Hobson, Charles J., and Frederick W. Gibson. "Policy Capturing as an Approach to Understanding and Improving Performance Appraisal: A Review of the Literatiure". Academy of Management Review (Vol. 8, No. 4, 1983): 640.

Hoffman, P. J. "The Paramorphic Representation of Clinical Judgment". Psychological Bulletin (March 1960): 116-31.

Hoffman, P. J., Slovic, P., and L. G. Rorer. "An Analysis-of Variance Model for the Assessment of Configural Cue Utilization in Clinical Judgment". Psychological Bulletin (May 1968): 338-349.

Hofstedt, T. R., and G. D. Hughes. "An Experimental Study of the Judgment Element in Disclosure Decisions". The Accounting Review (April 1977): 379-95.

Hofstedt, T. R., and J. C. Kinard. "A Strategy for Behavioral Accounting Research". The Accounting Review (January, 1970).

Ijiri, Yuji and Robert S. Kaplan. "The Four Objectives of Sampling in Auditing: Representative, Corrective, Protective, and Preventive". Management Accounting (December, 1970): 41-44.

Jensen, R. E. "An Experimental Design for Study of Effects of Accounting Variations in Decision Making". Journal of Accounting Research (Autumn 1966): 224-38.

Joyce, E. J. "Expert Judgment in Audit Program Planning". Supplement to the Journal of Accounting Research (1976): 29-60.

Joyce, E. J. and G. C. Biddle. "Are Auditors' Judgments Sufficiently Regressive"? Journal of Accounting Research (Autumn 1980).

Joyce, Edward J. and Robert Libby. "Behavioral Studies of Audit Decision Making". Journal of Accounting Literature (Vol. 1, 1982): 103.

Kahneman, D. "On the Psychology of Prediction". Psychological Review (July 1973): 237-51.

Kell, Walter G., and Richard E. Ziegler. Modern Auditing (New York: John Wiley & Sons, 1983).

Kennedy, H. A. "A Behavioral Study of the Usefulness of Four Financial Ratios". Journal of Accounting Research (Spring 1975): 97-116.

Keren, G. and R. Newman. "Additional Considerations with Regard to Multiple Regression and Equal Weighting". Organizational Behavior and Human Performance (1978) Vol. 22: 143-164.

Kida, Thomas. "An Investigation into Auditors' Continuity and Related Qualification Judgments". Journal of Accounting Research (Autumn 1980): 506-23.

Kinney, W. R. "A Decision Theory Approach to the Sampling Problem in Auditing". Journal of Accounting Research (Spring 1975): 117-32.

Kinney, W. R. "Decision Theory Aspects of Internal Control System Design/Compliance and Substantive Tests". Supplement to the Journal of Accounting Research (1975): 14-29.

Kinney, William R., Jr. and Wilfred C. Uecker. "Mitigating the Consequences of Anchoring in Auditor Judgments". The Accounting Review (January 1982): 55-69.

Libby, R. "Accounting Ratios and the Prediction of Failure: Some Behavioral Evidence". Journal of Accounting Research (Spring 1975): 150-61.

Libby, R. "The Use of Simulated Decision Makers in Information Evaluation". The Accounting Review (July 1975): 475-89.

Libby, R. "Man Versus Model of Man: Some Conflicting Evidence". Organizational Behavior and Human Performance (June 1976): 1-12.

Libby, R. "Man Versus Model of Man: The Need for a Nonlinear Model". Organizational Behavior and Human Performance (June 1976): 23-6.

Libby, R. "Bankers and Auditors' Perceptions of the Message Communicated by the Audit Report". Journal of Accounting Research (Spring 1979): 99-122.

Libby, R. Accounting and Human Information Processing: Theory and Applications (Prentice-Hall, 1981).

Loeb, G. The Battle for Investment Survival. (New York: Simon & Schuster, 1965).

Meigs, Walter B., Whittington, O. Ray, and Robert F. Meigs. Principles of Auditing, (Homewood, Illinois: Richard D. Irwin, 1982).

Messier, W. F. "An Analysis of Expert Judgment in the Materiality/Disclosure Decision". unpublished manuscript, University of Florida, 1979.

McClanahan, A. H. and L. K. Waters. "A Monte Carlo Investigation of Four Techniques for Capturing and Clustering Rater Strategies". Organizational Behavior and Human Performance (1976): 258-67.

Mock, T. J. and M. A. Vasarhelyi. "A Synthesis of the Informatioin Economics and Lens Models". Journal of Accounting Research (Autumn 1978): 414-23.

Moriarity, Shane. "Communicating Financial Information Through Multidimensional Graphics". Journal of Accounting Research (Spring 1979): 205-24.

Moriarity, S. "Judgment Based Definition of Materiality," Supplement to the Journal of Accounting Research (1979): 114-35.

Moriarity, S. and F. H. Barron. "Modeling the Materiality Judgments of Audit Partners". Journal of Accounting Research (Autumn 1976): 320-41.

Neter, J. and J. K. Loebbecke, Behavior of Major Statistical Estimators in Sampling Accounting Populations. Audit Reserarch Monograph No. 2 (New York: AICPA, 1975).

Pfaffenberger, Roger C. and James H. Patterson. Statistical Methods for Business and Econonmics. (Homewood, Illinois: Richard D. Irwin, 1981).

Reckers, P. M., and M. E. Taylor. "Consistency in Auditors' Evaluations of Internal Accounting Controls". Journal of Accounting, Auditing, and Finance (Fall 1979): 42-55.

Ricchiute, David N. Auditing: Concepts and Standards. (Cincinnati, Ohio: South-Western Publishing Co., 1982).

Robertson, J. C. and F. G. Davis. Auditing, (Plano, Texas: Business Publications, Inc., 1982): 361.

Schultz, Joseph J., Jr., and Sandra G. Gustavson. "Actuaries' Perceptions of Variables Affecting the Independent Auditor's Legal Liability". The Accounting Review (July 1978): 626-41.

Siegel, S. Nonparametric Statistics for the Behavioral Sciences (New York: McGraw-Hill, 1956).

Slovic, P. "Analyzing the Expert Judge: A Descriptive Study of a Stockbroker's Decision Processes". Journal of Applied Psychology (August 1969): 255-63.

Slovic, P., Fischhoff, B. and S. Lichtenstein. "Behavioral Decision Theory". Annual Review of Psychology (1977): 1-39.

Solomons, David. "The Historical Development of Costing." Studies in Costing (1952).

Summers, J. D., Taliaferro, J. D. and D. J. Fletcher. "Subjective Versus Objective Description of Judgment Policy" Psychonomic Science (February 1970): 249-50.

Thomas, W. C., and E. O. Henke. Auditing: Theory and Practice, (Boston, MA., Kent Publishing Co., 1983).

Ward, Bart H. "An Investigation of the Materiality Construct in Auditing". Journal of Accounting Research (Spring 1976): 138-52.

Weber, R. "Auditor Decision Making on Overall System Reliability: Accuracy, Consensus, and Usefulness of a Simulated Decision Aid". Journal of Accounting Research (Autumn 1978): 368-88.

Wright, William F. "Financial Information Processing Models: An Empirical Study". The Accounting Review, (July 1977): 676-89.

Zuber, George., Elliott, Robert., Kinney, William R., Jr., and James J. Leisenring. "Using Materiality in Audit Planning". Journal of Accountancy (March 1983): 46.

Appendix A

Fifty fictitious cases were developed for this study. A matrix of the cases' factor levels, 'mean estimation' statistical sample sizes, and the auditors' average recommended nonstatistical sample sizes is given in this matrix. For each case the auditors' mean recommended size is based on fifteen observations. The abbreviated column headings are as follows: OBS--case number; REC--auditors' average recommended nonstatistical sample sizes; MPU--'mean estimation' statistical sample size; SIGMA--population variance; UR--ultimate risk; IC--internal control risk; AR--analytical review risk; TD--risk of incorrect acceptance; TE--tolerable error; N--population size; E--expectation of error; ALPHA---risk of incorrect rejection; PAST--last year's nonstatistical sample size; ASSET--balance sheet value of account; and SALES--gross revenues.

TABLE 14

Matrix of Cases

SAS

OBS	REC	MPU	SIGMA	UR	IC	AR	TD	TE	N	E	ALPHA	PAST	ASSET	SALES
1	48	18	185	0.05	1.000	0.500	0.10	20000	150	6000	0.05	46	250000	2000000
2	40	10	135	0.05	1.000	0.500	0.10	20000	150	6000	0.05	2	250000	2000000
3	49	27	235	0.05	1.000	0.500	0.10	20000	150	6000	0.05	45	250000	2000000
4	57	14	185	0.05	1.000	0.500	0.10	20000	150	6000	0.10	54	250000	2000000
5	44	8	135	0.05	1.000	0.500	0.10	20000	150	6000	0.10	8	250000	2000000
6	43	22	235	0.05	1.000	0.500	0.10	20000	150	6000	0.10	12	250000	2000000
7	49	24	185	0.05	1.000	0.500	0.10	20000	150	6000	0.01	72	250000	2000000
8	43	14	135	0.05	1.000	0.500	0.10	20000	150	6000	0.01	6	250000	2000000
9	46	35	235	0.05	1.000	0.500	0.10	20000	150	6000	0.01	63	250000	2000000
10	48	28	200	0.05	1.000	0.500	0.10	20000	150	6000	0.01	28	250000	2000000
11	51	14	100	0.05	0.800	0.625	0.10	25000	250	7000	0.01	5	320000	3000000
12	50	48	120	0.05	0.800	0.625	0.10	15000	250	4000	0.01	67	180000	2000000
13	41	60	70	0.05	0.833	0.600	0.10	10000	375	2000	0.05	11	125000	3000000
14	42	22	50	0.05	0.833	0.600	0.10	7500	250	2000	0.10	12	90000	1000000
15	57	31	60	0.05	0.833	0.600	0.10	10000	300	3000	0.05	71	125000	2000000
16	41	74	100	0.05	0.500	0.500	0.20	15000	500	4500	0.05	65	180000	2000000
17	65	44	125	0.05	1.000	1.000	0.05	10000	150	3300	0.01	34	125000	2000000
18	50	97	150	0.05	0.800	0.625	0.10	20000	500	6000	0.10	18	250000	2000000
19	54	35	100	0.05	0.800	0.625	0.10	25000	400	7500	0.01	25	320000	3000000
20	52	23	80	0.05	0.800	0.625	0.10	25000	400	7000	0.01	23	320000	3000000
21	67	25	250	0.05	1.000	1.000	0.05	50000	250	15000	0.01	83	360000	2000000
22	58	17	200	0.05	0.500	1.000	0.10	30000	200	10000	0.05	49	450000	2000000
23	74	55	175	0.05	1.000	0.500	0.10	20000	250	6000	0.01	53	300000	2000000
24	48	15	200	0.05	0.833	0.600	0.10	50000	260	15000	0.01	15	450000	3000000
25	40	59	50	0.05	0.800	0.625	0.10	3000	175	1000	0.05	7	50000	2000000
26	32	17	100	0.01	0.400	0.500	0.05	10000	125	3000	0.05	17	125000	2000000
27	61	9	50	0.01	1.000	1.000	0.01	10000	130	3000	0.01	39	130000	2000000
28	45	33	50	0.01	0.200	1.000	0.05	20000	250	5000	0.10	45	200000	2000000
29	35	37	105	0.01	1.000	1.000	0.01	5000	80	1500	0.01	8	70000	2000000
30	58	33	75	0.01	1.000	1.000	0.01	15000	250	4800	0.01	43	190000	2000000
31	57	20	70	0.05	1.000	1.000	0.05	5000	100	1500	0.05	61	80000	2000000
32	63	21	100	0.05	1.000	0.500	0.10	20000	250	4000	0.01	80	190000	2000000
33	38	7	105	0.01	1.000	1.000	0.01	15000	80	4800	0.01	26	180000	2000000
34	59	27	150	0.05	1.000	0.500	0.10	20000	250	6000	0.10	28	285000	2000000
35	31	15	100	0.01	0.400	0.500	0.05	10000	125	3000	0.10	5	120000	2000000
36	42	35	50	0.05	0.833	0.600	0.10	7500	250	2000	0.01	6	125000	2000000
37	51	20	100	0.05	0.500	0.500	0.20	15000	250	4500	0.05	47	250000	3000000
38	90	17	100	0.05	1.000	1.000	0.05	25000	250	8000	0.01	57	400000	3000000
39	46	93	95	0.05	0.833	0.600	0.10	7500	250	2000	0.01	24	150000	2000000
40	67	33	50	0.01	1.000	1.000	0.01	25000	600	7000	0.01	33	300000	3000000
41	44	27	100	0.20	1.000	1.000	0.20	10000	200	3000	0.05	7	150000	2000000
42	29	15	150	0.05	0.300	0.500	0.33	15000	170	4800	0.05	15	250000	2000000
43	43	48	120	0.05	0.500	0.500	0.20	8000	190	2000	0.05	28	115000	2000000
44	32	43	150	0.05	0.300	0.500	0.33	8000	170	2500	0.05	43	120000	2000000
45	68	15	60	0.10	1.000	0.500	0.20	8000	190	2500	0.05	75	125000	2000000
46	50	19	150	0.05	0.500	0.500	0.20	25000	300	5000	0.10	0	300000	2000000
47	71	30	150	0.10	1.000	1.000	0.10	12000	150	3900	0.05	10	180000	2000000
48	28	25	180	0.05	0.300	0.500	0.33	7500	100	3000	0.05	5	125000	2000000
49	30	8	50	0.05	0.500	0.500	0.20	5000	90	1500	0.01	8	100000	2000000
50	29	11	150	0.05	0.500	0.500	0.20	17500	165	5000	0.10	0	180000	2000000

The vita has been removed from the scanned document

Modeling Auditor Judgment in Nonstatistical Sampling

by

William J. Read

(ABSTRACT)

Since its issuance in June 1981, Statement on Auditing Standards (SAS) No. 39, "Audit Sampling," has been the center of much controversy. Practitioners are voicing their concerns as they anticipate difficulties in designing, selecting, and evaluating a nonstatistical sampling procedure in accordance with SAS 39.

This proposed exploratory study seeks to identify those factors that underlie the auditor's judgment with respect to nonstatistical sample size decisions in substantive tests. The research will utilize Egon Brunswik's Lens Model to provide mathematical representations of the auditor's judgment process. Correlational statistics will be used to assess judgment accuracy, agreement (consensus), and auditor "self-insight" into his decision process.

The study will provide empirical insight, into whether the auditor's determination of the appropriate extent of testing is consistent with his judgment as to the assurance level needed from his sampling application, or conversely, the degree of risk he is willing to accept. The ability of auditors to formulate their sample size decisions properly

is crucial because of their impact on audit effectiveness and efficiency. In addition, this project should provide additional evidence bearing upon the arguments of both proponents and opponents of SAS 39.