1979


Johng Yul Lee
Louisiana State University and Agricultural & Mechanical College

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_disstheses

Recommended Citation
https://digitalcommons.lsu.edu/gradschool_disstheses/3339
INFORMATION TO USERS

This was produced from a copy of a document sent to us for microfilming. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help you understand markings or notations which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure you of complete continuity.

2. When an image on the film is obliterated with a round black mark it is an indication that the film inspector noticed either blurred copy because of movement during exposure, or duplicate copy. Unless we meant to delete copyrighted materials that should not have been filmed, you will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed the photographer has followed a definite method in "sectioning" the material. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For any illustrations that cannot be reproduced satisfactorily by xerography, photographic prints can be purchased at additional cost and tipped into your xerographic copy. Requests can be made to our Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases we have filmed the best available copy.
LEE, JOHNG YUL
A TEST OF ALTERNATIVE PRIOR PROBABILITY ELICITATION METHODS IN ASSESSING THE RELIABILITY OF INTERNAL CONTROL SYSTEMS FOR AUDIT DECISIONS.

THE LOUISIANA STATE UNIVERSITY AND AGRICULTURAL AND MECHANICAL COL., PH.D., 1979

University Microfilms International 300 N. ZEEB ROAD, ANN ARBOR, MI 48106
PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark □.

1. Glossy photographs □
2. Colored illustrations □
3. Photographs with dark background □
4. Illustrations are poor copy □
5. Print shows through as there is text on both sides of page □
6. Indistinct, broken or small print on several pages □ throughout
7. Tightly bound copy with print lost in spine □
8. Computer printout pages with indistinct print □
9. Page(s) □ lacking when material received, and not available from school or author □
10. Page(s) □ seem to be missing in numbering only as text follows □
11. Poor carbon copy □
12. Not original copy, several pages with blurred type □
13. Appendix pages are poor copy □
14. Original copy with light type □
15. Curling and wrinkled pages □
16. Other □

University Microfilms International
300 N. ZEEB RD., ANN ARBOR, MI 48106 (313) 761-4700
A TEST OF ALTERNATIVE
PRIOR PROBABILITY ELICITATION METHODS
IN ASSESSING THE RELIABILITY OF
INTERNAL CONTROL SYSTEMS FOR AUDIT DECISIONS

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Accounting

by

Johng Yul Lee
B.S., Seoul National University, 1970
M.S., Louisiana State University, 1977
May, 1979
ACKNOWLEDGEMENTS

I wish to express my appreciation and indebtedness to all of the individuals and organizations which have provided invaluable assistance to this dissertation.

My sincere appreciation is extended to my committee, Professors Alan J. Winters (Chairman), Vincent C. Brenner, C. Willard Elliott, Bart P. Hartman, and Don L. Marx, for their full cooperation. Their comments and criticisms were most constructive and helpful.

To Professor Robert L. Winkler of Indiana University, I convey my gratitude for his helpful advice on the methodologies used in this research. My gratitude is also extended to Professors Paul R. Lohnes and Ronald J. Huefner of State University of New York at Buffalo for their help in the survey and analysis of the results.

To my mother this dissertation is dedicated in hopes that it can express my deep appreciation for her sacrifice and love which has made me what I am today. To my wife Jungmi, a great typist, I express my appreciation for her enormous help and indulgence in my doctoral study.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>ii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vi</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>vii</td>
</tr>
<tr>
<td>Chapter</td>
<td></td>
</tr>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Purpose of the Inquiry</td>
<td>1</td>
</tr>
<tr>
<td>Hypotheses</td>
<td>3</td>
</tr>
<tr>
<td>Methodology of this Study</td>
<td>4</td>
</tr>
<tr>
<td>Scope of the Study</td>
<td>6</td>
</tr>
<tr>
<td>Terminology</td>
<td>7</td>
</tr>
<tr>
<td>Organization of the Contents</td>
<td>9</td>
</tr>
<tr>
<td>II. THE AUDIT PROCESS AND THE LEVELS OF ASSURANCE</td>
<td>12</td>
</tr>
<tr>
<td>Audit Satisfaction</td>
<td>12</td>
</tr>
<tr>
<td>Internal Control and Compliance</td>
<td>15</td>
</tr>
<tr>
<td>Substantive Tests</td>
<td>17</td>
</tr>
<tr>
<td>III. BAYESIAN APPROACH TO STATISTICAL INFERENCE IN AUDITING</td>
<td>20</td>
</tr>
<tr>
<td>Probabilistic Evaluation of Internal Control</td>
<td>20</td>
</tr>
<tr>
<td>Bayesian Inference</td>
<td>21</td>
</tr>
<tr>
<td>Bayesian Method in Auditing</td>
<td>22</td>
</tr>
<tr>
<td>conjugate Prior Distributions</td>
<td>26</td>
</tr>
<tr>
<td>IV. THE ASSESSMENT OF PRIOR PROBABILITY DISTRIBUTIONS</td>
<td>29</td>
</tr>
<tr>
<td>Subjective Probability Assessment Methods</td>
<td>29</td>
</tr>
<tr>
<td>Probability Density Function (PDF)</td>
<td>31</td>
</tr>
<tr>
<td>Cumulative Density Function (CDF)</td>
<td>33</td>
</tr>
<tr>
<td>Psychometric Ranking</td>
<td>35</td>
</tr>
<tr>
<td>Hypothetical Future Samples (HFS)</td>
<td>36</td>
</tr>
<tr>
<td>Chapter</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Equivalent Prior Sample (EPS)</td>
<td>38</td>
</tr>
<tr>
<td>Ratio Estimation</td>
<td>38</td>
</tr>
<tr>
<td>Direct Specification of Parameters</td>
<td>38</td>
</tr>
<tr>
<td>Credibility Forecast Diagram (Potential Surprise Concept)</td>
<td>39</td>
</tr>
<tr>
<td>Range Betting Method</td>
<td>40</td>
</tr>
<tr>
<td>Empirical Tests of Various Methods</td>
<td>41</td>
</tr>
<tr>
<td>Winkler's Study</td>
<td>42</td>
</tr>
<tr>
<td>Corless's Study</td>
<td>45</td>
</tr>
<tr>
<td>Felix's Study</td>
<td>47</td>
</tr>
<tr>
<td>Evaluation of &quot;the Four Methods&quot;</td>
<td>52</td>
</tr>
<tr>
<td>CDF</td>
<td>53</td>
</tr>
<tr>
<td>PDF</td>
<td>54</td>
</tr>
<tr>
<td>EPS</td>
<td>54</td>
</tr>
<tr>
<td>V. THE PURPOSE OF THE STUDY AND THE METHODOLOGY</td>
<td>59</td>
</tr>
<tr>
<td>The Purpose</td>
<td>59</td>
</tr>
<tr>
<td>The Methodology</td>
<td>65</td>
</tr>
<tr>
<td>VI. ANALYSIS OF THE RESULTS</td>
<td>75</td>
</tr>
<tr>
<td>The Auditor's Perception of Hypothetical Samples Techniques</td>
<td>75</td>
</tr>
<tr>
<td>The Consistency in SPD Assessments</td>
<td>84</td>
</tr>
<tr>
<td>The Effect of Background Factors on Discrepancies</td>
<td>93</td>
</tr>
<tr>
<td>Compatibility of HFS with EPS</td>
<td>98</td>
</tr>
<tr>
<td>VII. SUMMARY, CONCLUSIONS, AND SUGGESTED FUTURE RESEARCH</td>
<td>101</td>
</tr>
<tr>
<td>Objectives Restated</td>
<td>101</td>
</tr>
<tr>
<td>Summary and Conclusions</td>
<td>102</td>
</tr>
<tr>
<td>Suggested Future Research</td>
<td>108</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>113</td>
</tr>
<tr>
<td>APPENDIX A - THE QUESTIONNAIRE</td>
<td>117</td>
</tr>
<tr>
<td>APPENDIX B - COMMENTS ON FELIX'S ARTICLE</td>
<td>123</td>
</tr>
<tr>
<td>APPENDIX C - FELIX'S RESPONSE</td>
<td>130</td>
</tr>
<tr>
<td>VITA</td>
<td>133</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Mean Ratings of the Techniques on the Clarity Scale (The Auditor Group)</td>
<td>76</td>
</tr>
<tr>
<td>2. $k \times n$ Table of Ranks ($k=36$ and $n=3$) for Friedman Test</td>
<td>79</td>
</tr>
<tr>
<td>3. Mean Ratings of the Techniques on the Clarity Scale (The Student Group)</td>
<td>83</td>
</tr>
<tr>
<td>4. Discrepancies between Quartiles of Two Prior Distributions (the EPS and Discrete PDF Methods)</td>
<td>87</td>
</tr>
<tr>
<td>5. Discrepancies between Quartiles of Two Prior Distributions (the HFS and Discrete PDF Methods)</td>
<td>88</td>
</tr>
<tr>
<td>6. Median Responses Falling within Each Error Rate Interval</td>
<td>92</td>
</tr>
<tr>
<td>7. Results of Simple Bivariate Regressions of Backgrounds on Discrepancies between Median Responses (Case A)</td>
<td>94</td>
</tr>
<tr>
<td>8. Results of Multiple Regressions of Backgrounds on Discrepancies between Median Responses (Case A)</td>
<td>95</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure                                         Page
1. A Typical Probability Distribution Curve     32
2. Graphs of Cumulative Distribution Functions  91
    for Subject 8
ABSTRACT

The second standard of field work requires that the auditor study and evaluate the client's system of internal control to determine the extent of substantive tests for a certain desired level of assurance. To use the formula for the reliability level for substantive tests set forth in Statements on Auditing Standards, the reliance assigned to internal control should be expressed by the auditor in the form of probability. For application of the Bayesian method in statistical sampling and the decision theoretic approaches to decision making in auditing, the auditor's subjective evaluation of internal control should be expressed in the form of a probability distribution.

Auditors, however, are generally not statisticians, and thus a need arises for an appropriate method of expressing their subjective evaluation. Therefore, study of the appropriate subjective probability assessment techniques, which would not overwhelm the practicing auditor's relatively modest background in statistics and probability is essential.

The purpose of this study was to test the subjective probability elicitation techniques which seem to be most
congruent with the audit environment and the auditor's background, and to investigate the basic qualifications required for the auditor to be able to assess his (her) subjective probability distributions (SPD) in a reasonably consistent manner.

The following hypotheses existed in the study:
(1) There are certain methods of eliciting subjective probability which are most appealing to auditors.
(2) These methods, called hypothetical samples techniques, are also very easy to apply by auditors in the evaluation of internal control for audit decision making. (3) The use of these hypothetical samples techniques will increase consistency in subjective probability assessments in the auditing context. (4) Personal backgrounds of auditors should have some impact on the discrepancies in assessed distributions. (5) There may be basic, minimum qualifications of auditors for reasonably reliable assessments. (6) There may be a combination of two or more elicitation methods compatible with each other, which will insure a very high quality probability assessment.

The research involved a literature survey, administration of questionnaires to experienced auditors, and analysis of the survey results.
The study confirmed the appealing nature of the two hypothetical samples techniques - the equivalent prior samples method and the hypothetical future samples method. Audit experience played a major role in making the two hypothetical samples techniques appealing to the auditors. The use of the hypothetical samples techniques was found to increase consistency in subjective probability assessments in the auditing environment. The research also indicated that auditors feel more comfortable with point estimation than with distribution estimation. Auditors might, it was found, have a reasonably uniform range of error expectations for a given circumstance.

There was no single personal background factor of auditors which accounted for most of the variances in the discrepancies in subjective probability assessments. Slightly negative relationships were found between each personal background factor and the discrepancy in assessment, indicating that auditing experience and sampling knowledge enhance consistency in SPD assessment. The size of the majority of the auditor's clients had virtually no impact on the resulting consistency.

One conclusion reached based on the experiment was that the auditors could use both the equivalent prior samples method and the hypothetical future samples method, check resulting discrepancies at each beta quartile value,
and compare them to the average discrepancies reported in this study to check the magnitude of the resulting discrepancies. The compatibility of the two methods with each other was indicated. Except suggesting that a reasonable amount of audit experience and some background in statistics help auditors to obtain reasonably consistent SPD assessments, the study did not provide the minimum qualifications required of an auditor for an appropriate quality assessment.
CHAPTER I
INTRODUCTION

Purpose of the Inquiry.

The second standard of field work requires that the auditor study and evaluate the client's system of internal control to determine the extent of substantive tests necessary for a certain desired level of assurance.\(^1\)

To use the formula for the reliability level applicable to substantive tests given in Statement on Auditing Standards, the reliance assigned to internal control should be expressed by the auditor in the form of probability.\(^2\)

In statistical sampling for both tests of transactions and substantive tests, the sample size could be significantly decreased while maintaining the same level of reliability by using another statistical technique called the Bayesian method.

For the application of these techniques, the auditor's subjective evaluation of internal control should be expressed in the form of a probability distribution.

\(^1\)American Institute of Certified Public Accountants, Statement on Auditing Standards, 1973, p. 5.

\(^2\)SAS 1, p. 53.
If a person who assigns subjective probabilities adopts certain principles to ensure consistent assignments, it is possible to prove that this probability measure satisfies the requirements of probability set function. The problem a researcher in subjective probability encounters first in the study is in the abundance of the literature. Numerous techniques in the assessment of subjective probability have been attempted by many scholars. As long as the technique for the quantification of judgments satisfies the probability axioms, any technique could possibly be tried for an experiment.

Of the numerous techniques, some of the methods are relevant to particular situations, which means that the methods are not generally applicable to other cases. A substantial portion of the experiments used naive subjects, mostly psychology students taking courses or participating in laboratory experiments, and the results of these experiments cannot be applied to auditing in which only experienced or well informed auditors make decisions or estimate risk levels.

In practice, when the probability assessor is not a statistician, as is the case in the auditing environment, the more elegant and sophisticated theoretical suggestions or methods cannot be used. If the assessor is not thoroughly familiar with the formal notions of probability, those sophisticated methods of
eliciting subjective probability would not be comprehensible to him. Auditors are not generally statisticians, and there should be an appropriate method(s) of expressing the subjective evaluation of qualitative audit evidence for auditors. Therefore, study of the appropriate subjective probability assessment technique(s), which would not "overwhelm" the practicing auditor's relatively modest background in statistics and probability, is essential.

The purpose of this dissertation is to test the subjective probability elicitation techniques which seem to be most congruent with the audit environment and auditors' backgrounds, and to investigate the basic qualifications which are necessary for auditors to be able to assess their subjective probability distributions in a reasonably consistent manner.

Hypotheses

Preliminary research into psychological, statistical, mathematical, and accounting literature indicated that certain viable hypotheses existed that could be tested in this study. These include the following:

1. There are certain methods of eliciting subjective probability which are most appealing to auditors.
2. These methods, called hypothetical samples techniques, are also very easy to apply for auditors in the evaluation of internal control for audit decision making.

3. The use of these hypothetical samples techniques will increase the consistency in subjective probability assessments in the auditing context.

4. Personal backgrounds of auditors should have some impact on the discrepancies in assessed subjective probability distributions.

5. There may be basic, minimum qualifications of auditors required for reasonably reliable assessments of subjective probability distributions.

6. There may be a combination of two or more subjective probability elicitation methods compatible with each other, which will insure a very high quality probability assessment.

After these hypotheses were established from the preliminary research, methodological plans were formulated.

Methodology of this Study

The general methodology applied in this study includes the following:

Phase 1: Literature survey. An extensive survey was made of the literature in psychology, probability and statistics, mathematics, systems, and
accounting. The reliability assessment was related to the audit process and the levels of assurance in auditing. Subsequently, the Bayesian approach to statistical inference in auditing was discussed. A comprehensive review of subjective probability elicitation techniques was made in the final phase of the literature survey and the implication of the various methods in the audit environment was analyzed while previous applications of these methods to auditing problems were discussed.

**Phase 2: Case development and questionnaire design.** An audit case was developed for the experiment and a questionnaire was designed to test the hypothetical samples techniques in comparison with the discrete probability distribution function technique. In designing the questionnaire, the previous interview-format questionnaires used by other experimenters served as the initial framework. After the initial format of the questionnaire was designed, the appropriateness and usefulness of it was tested in two review sessions.

**Phase 3: Actual conduct of survey.** Questionnaires were administered to the auditors of major accounting firms in the Fall of 1978. Then managing partner of each participating firm was provided with multiple copies of the questionnaire and return envelopes, and administered these questionnaires to their audit staff which returned the responses, with
their identities specified, directly to the experimenter. Due to the close relationship existing between the accounting educators and the firms located in the area, the response rate was very high. More detailed information on the survey is provided in Chapter Five. A copy of the questionnaire is presented in Appendix A.

**Phase 4: Analysis of survey results.** The responses obtained were analyzed through parametric and nonparametric statistical tests. The various statistical analyses were made for each specific purpose identified in the objectives of the study, and for each individual methodology. The results of the statistical analyses and interpretations of the results are presented in Chapter Six.

**Scope of the Study**

Some experimental limitations restricted the scope of the study. Most of the limitations were due to the inherent nature of an experiment such as this; budgetary constraints, limited time period available for the study, difficulty in maintaining the interest of the subject-auditors for an extended period of time, and difficulty in making the same subject available for subsequent phases of the study.

In addition to the limitations related to the subjects, an inherent limitation was involved in the
case used in the experiment. The background of the business, history of the company, organizational details, existing internal control system, and the results of the past audit results including sampling results could not be fully explained to the subjects because of the constraints discussed above. Accordingly, it appeared that the subject-auditors interpreted the facts of the case differently. Therefore, it is possible that the results observed reflect to some extent this bias in the case interpretation.

More detailed discussion of the implications of the limitations in the experiment is presented in Chapter Six, and some suggestions to alleviate the deficiencies experienced in this study for future research in this context are made in Chapter Seven.

Terminology

In order to promote a meaningful interpretation of the discussions made in this study, the following terminology description is provided:

1. ASSURANCE refers to the confidence that the auditor has of the fair presentation of the financial statements.

2. ATTRIBUTE refers to quality characteristics of transaction data input and to quality characteristics of processed data output for a system.
3. AUDIT refers to the process of examining financial statements for the purpose of rendering an informed opinion, and this study is restricted to the independent audit function as compared to internal management audit function.

4. AUDITOR refers to the independent certified public accountant or member of an accounting firm performing an audit.

5. CLIENT refers to the entity which is audited by an auditor.

6. CPA refers to a certified public accountant.

7. CREDIBLE INTERVAL refers to the interval of probability that is centered at a certain point estimate of a probability.

8. CUMULATIVE DISTRIBUTION FUNCTION refers to the function relating the various values of a random variable to the corresponding cumulative probabilities.

9. F(x) represents the cumulative distribution function of a random variable x.

10. ELICITATION refers to assessment and is used only in the discussion of probability.

11. INTERNAL CONTROL refers to the means by which assets are safeguarded, the accuracy and reliability of accounting data are checked, operational efficiency is promoted, and adherence to prescribed organizational policies are encouraged.
12. PRIOR DISTRIBUTIONS refers to prior probability distributions.

13. PRIOR PROBABILITY DISTRIBUTIONS refers to a probability distribution consisting of a set of probabilities which summarizes the information concerning a parameter of a random variable in Bayesian theorem.

14. RELIABILITY refers to the confidence level the auditor has of a proposition or item.

15. SUBJECTIVE PROBABILITY refers to a measure of the confidence which a particular individual has in the truth of a particular proposition.

Organization of the Contents

Chapter II contains the discussion on the levels of assurance as related to each phase of the audit process. An attempt was made to identify the objective of measuring the reliability level of an internal control system within the framework of the overall audit process and evidence gathering.

After a discussion on the probabilistic evaluation of internal control, the Bayesian approach to statistical inference in auditing was introduced in Chapter III. The concepts and procedures unique to the Bayesian approach in auditing were reviewed and the significance of prior probability distributions for this methodology
was discussed with an emphasis on conjugate prior distributions which provides helpful working definitions to the Bayesian approach.

Chapter IV includes the discussion on the concepts and procedures of various prior probability assessment methods. A comprehensive review of concepts and procedures was followed by a feasibility analysis of alternative methods based on the results of empirical tests using the methods presented. More in-depth evaluation was made of the methods which seemed to be compatible with the auditing environment. The peculiar nature of the audit and the unique background of the auditor were related to each characteristic of the methods presented.

The purposes of the study and the methodology were discussed in Chapter V. Specific purposes were defined, and for each purpose appropriate methodology was applied. The methodologies include parametric and nonparametric statistical analyses and other analytical techniques.

Analysis of the results of the study was presented in Chapter VI. The Statistical Package for the Social Sciences was used for the necessary computer analysis of the data collected. Selected key parameters and scores were tabulated in proper forms, and presented in the chapter.
Chapter VII contains the summary of the findings and the conclusions drawn from the study. Suggestions for future research based on the experience acquired from this experiment were presented in the last section of this chapter.
CHAPTER II

THE AUDIT PROCESS AND THE LEVELS OF ASSURANCE

Audit Satisfaction

The term, audit satisfaction, has been understood to mean that the reliability of financial statements which the auditor assesses should be higher than a "reasonable level" for a warranted assertion. Based on this unspecified "reasonable level" of reliability, the reliabilities required from substantive audit tests are derived by the auditor. To justify an assertion of a proposition, that assertion should be proved to be true. When an assertion is made by a professional, the assertion is warranted. When the professional warrants an assertion, he must have formed a belief to support the warranted assertion and this is generally termed assurance.

This warranted assertion is distinguished from positive knowledge because of its lack of absolute certainty. The degree of certainty ranges from certain truth to certain falsehood. The Report of the Committee on Basic Auditing Concepts of the American Accounting

Association (1969-1971) refers to the belief and the degree of credibility.

Ideally, an investigator should not express his belief concerning a proposition without expressing the degree of credibility attached to the proposition being judged.\(^4\)

Thus, an auditor assigns a degree of credibility to the belief formed on each proposition under consideration. He will, then, assess the weight to be given to each proposition based on the importance of the amount involved, the effect on other accounts, and so forth. This weight is symbolized as \(w_i\) which can be scaled to meet the conditions:

\[
0 < w_i < 1, \quad \text{and} \quad \sum_{i=1}^{n} w_i = 1
\]

\((n: \text{number of propositions})\)

Then the total level of assurance \((T)\) for all the propositions, \(p_i's\), could be expressed as

\[
T = \sum_{i=1}^{n} w_i r_i \quad \text{and} \quad 0 \leq T \leq 1,
\]

where \(r_i\) represents the degree of credibility for each proposition subject to

\[
0 \leq r_i \leq 1
\]

\(^4\text{Report of the Committee on Basic Auditing Concepts, 1969-71, p. 41.}\)
The above multiplicative model is not suggested strongly here as a normative scheme for the aggregation of all individual levels of credibility for all the propositions in the financial statements of a client. Rather, this multiplicative model gives a reasonably conservative expression of the overall level of assurance decision process given the accounting profession's limited knowledge of the effect of interconnected propositions on the overall level of assurance.

The auditor must gather sufficient evidence to support a belief on each proposition. The degree of credibility for a proposition is composed of three elements; (1) the degree of credibility for each account or area of the system of internal control, denoted as $c_i$, (2) the subjective judgment or the statistical inference made from the substantive tests, $s_i$, and (3) the effect of interconnected propositions on the particular proposition. The discussion on the last element is beyond the scope of this study and I will confine the study to $c_i$ and $s_i$. In a functional form the relationship of $c_i$ and $s_i$ is indicated as

$$r_i = f(c_i, s_i) \quad i = 1, \ldots, n$$

where

$$0 \leq c_i \leq 1$$

$$0 \leq s_i \leq 1$$
If the system review and the compliance tests of internal control reveal an extremely high degree of credibility, that is if $c_i$ is almost equal to 1, then the substantive test could be made at a low level, and still the auditor would have a high degree of credibility for the proposition under consideration, assuming that the other audited accounts do not indicate an effect of interconnected propositions on the proposition $P_i$, and that there is no management override resulting in unanticipated errors which might have passed through the system undetected.

Internal Control and Compliance

The audit process begins with the evaluation of the client's system of internal control. For each class of transaction and/or system segment an examination is made as to the types of likely errors and the existence of the control procedures which should prevent and detect these errors. The importance of the study and evaluation of internal control in auditing is documented by the second standard of field work, which states:

There is to be a proper study and evaluation of the existing internal control as a basis for reliance thereon and for the determination of the resultant extent of
the tests to which auditing procedures are to be restricted.\textsuperscript{5}

Cushing defined the objective of examining internal control as the determination of the likelihood of specified types of errors being able to get through the system undetected.\textsuperscript{6} Morarity further described this:

> The auditor has learned, through experience, the types of errors which can be expected. The auditor also has learned that specific controls can be instituted to detect and correct these errors. Thus, if one auditor notices that a specific control is missing in the client's system, the auditor concludes that there is a positive probability that an anticipated type of error may exist, uncorrected, in the client's statements.\textsuperscript{7}

The next step is to test whether the prescribed procedures were applied properly by appropriate personnel. The results of this compliance test are used in determining the nature and the extent of substantive tests. Especially for those areas with weak controls and compliance, extended tests will be performed to detect the


"anticipated type of errors." Tests of compliance are not necessary if the control procedures are not to be relied upon to determine the extent of substantive tests. Such cases arise where the audit effort required to test compliance with the procedures to justify reliance on them in making substantive tests would exceed the reduction in effort that could be achieved by such reliance.\footnote{AICPA, \textit{Statement on Auditing Standards}, p. 28.}

In practice, the same sample may be taken for dual-purpose tests; compliance tests and substantive tests. Subjective evaluation of tests of compliance is possible. However, quantitative evaluation based on statistical sampling should be more useful in determining sample size and in evaluating sample results.\footnote{AICPA, \textit{Statement on Auditing Standards}, p. 30.}

**Substantive Tests**

Substantive tests are necessary to reduce the risk that any material errors will exist in the financial statements undetected by internal control. To ascertain that the risk of such errors is at a reasonably low level, the auditor tests transactions and/or account balances...
based on subjective judgments or using statistical sampling. Of course, evaluation of the sampling results is made as to both the precision of the range of values and reliability (confidence level) based on the auditor's judgment. The reliability level of substantive tests varies inversely with the reliability of internal control as measured by the auditor. The AICPA's Statement on Auditing Standards gives a model for deriving the reliability level for substantive tests developed from established probability theorems:

\[
S = 1 - \frac{1 - R}{1 - C}
\]

where \( S \) = Reliability level for substantive tests.

\( R \) = Combined reliability level desired.

\( C \) = Reliance assigned to internal accounting control and other relevant factors.\(^{10}\)

The relationship between \( C \), \( S \), and \( R \) can be shown more clearly by transforming the above equation. After a simple transformation the above equation becomes

\[
R = C + S - CS,
\]

or

\[
R = C + (1 - C)S
\]

For example, the following selected values of \( C \) and \( S \) are related to each other at the combined reliability

\(^{10}\)AICPA, *Statement on Auditing Standards*, p. 53.
level of 95 per cent, and illustrate the inverse relationship of the reliability of internal control systems and the extent of substantive tests.

<table>
<thead>
<tr>
<th>R</th>
<th>C</th>
<th>1 - C</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>95%</td>
<td>5%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>90%</td>
<td>10%</td>
<td>50%</td>
<td></td>
</tr>
<tr>
<td>95%</td>
<td>70%</td>
<td>30%</td>
<td>83%</td>
</tr>
<tr>
<td>50%</td>
<td>50%</td>
<td>90%</td>
<td></td>
</tr>
<tr>
<td>30%</td>
<td>70%</td>
<td>93%</td>
<td></td>
</tr>
</tbody>
</table>

Statements on Auditing Standards suggests that the relative weight to be given to the respective sources of reliance (internal control, tests of details and other auditing procedures) are matters for the auditor's judgment in the circumstances. It is important for the discussion in subsequent chapters to note that the AICPA is assuming that the auditor has already quantified C to determine the extent of substantive tests.

---

11AICPA, Statement on Auditing Standards, p. 34.
CHAPTER III

BAYESIAN APPROACH TO STATISTICAL INFERENCE IN AUDITING

Probabilistic Evaluation of Internal Control

Recently, there have been some notable contributions to the measurement of the reliability of internal control systems. Yu and Neter\textsuperscript{12} used a stochastic model for this, Cushing\textsuperscript{13,14} and Bodnar\textsuperscript{15} employed reliability theory for mathematical analysis and design of internal control system, and Kinney\textsuperscript{16} explored the decision theory aspects of internal control system design, compliance and substantive tests.


Kinney's discussion, however, is based on the assumption that the auditor can express his knowledge of the internal control evaluation and compliance assessments in the form of a probability distribution over the two decision-relevant states; true state and state with errors. The approaches of Yu and Neter, and Cushing suggest the feasibility of measuring the reliability of internal control in mathematical probabilities. Yet the applicability of these techniques to actual auditing situations is yet to be tested in settings where the numerous assumptions they made are relaxed.

Bayesian Inference

Contrary to the classical approach to statistical inference which is based solely on sample evidence, the Bayesian approach formally utilizes information other than sample information. The Bayesian combines sample information with other available information using Bayes' theorem, and the resulting combination of information is the basis for inferential and decision-making procedures.

In making inferences about a parameter \( \theta \), if we can assume that the information concerning \( \theta \) can be summarized by a probability distribution consisting of a set of probabilities \( P(\beta = \theta_i) \) for \( i = 1, \ldots, J \), and based on another assumption that \( \theta \) can only take on \( J \)
possible values $\theta_1, \ldots, \theta_j$, this probability distribution is called the prior distribution of $\theta$.

The outcome of the sample can be expressed in terms of a likelihood function which can be combined with the prior distribution to arrive at a combined distribution called the posterior distribution.

The motivation for Bayesian methods is essentially the desire to base inferences and decisions on any and all available information, regardless of the nature of the information.

**Bayesian Method in Auditing**

Because of the particular nature of statistical inference in auditing, the Bayesian approach has been advocated by many authors. Tracy, Sorensen, Kraft,


18 Winkler and Hays, p. 473.


Slotnick, Wilson, Birnberg, Knoblett, and Corless have each addressed various applications of Bayesian statistics to auditing. In auditing, the auditor has two different kinds of information; the subjective evidence he has collected (unquantified C in Chapter II), and the statistical sampling results (quantified). This provides an appropriate setting for Bayesian analysis. To apply Bayesian techniques, the auditor subjectively evaluates the strength of internal control, results of related audit tests, results of past audit examinations and reputation of the client; and expresses his belief about the audit population as a prior probability distribution.


h(θ) of the random variable θ that represents the error rate in the accounts and has a probability distribution over the set Ω. A likelihood function k(y|θ) is then obtained by statistically evaluating the sample result. y is a value of the statistic Y which is a function of a random sample whose distribution depends upon θ, a random determination of the random variable θ.

We can find the probability distribution function (pdf) of Y for every given θ; that is, we can find the conditional pdf of Y, given θ = θ, which is the above likelihood function k(y|θ). Using Bayes' formula, the prior distribution h(θ) is mathematically combined with the likelihood function h(θ) into a posterior probability distribution f(θ|y) as follows:

$$f(\theta|y) = \frac{h(\theta)k(y|\theta)}{\int h(\theta)k(y|\theta)d\theta}$$  \hspace{2cm} (Continuous case)

$$P(\theta_i|y) = \frac{P(\theta_i)P(y|\theta_i)}{\sum P(\theta_i)P(y|\theta_i)}$$  \hspace{2cm} (Discrete case)

This posterior distribution, f(θ|y) or P(θ_i|y), may be regarded as a reflection of the meaning of all audit evidence (both sampling and nonsampling) taken as a whole.²⁷

²⁷Corless, p. 557.
In classical statistics, all inferences are based solely on the sampling results. Some classical techniques are based directly upon the likelihood function as can be seen in maximum likelihood estimators. For any $\theta_i$, $P(y|\theta_i)$ can be thought of as the likelihood of the sample result $y$, given $\theta_i$. The likelihood function is uniquely determined according to the underlying distributional form of the sample, such as a binomial distribution for the account error rates in attribute sampling.

When graphically illustrated, the rather diffuse distribution of the priors in ordinary cases would change to a posterior distribution which would be concentrated around a certain error rate. Thus the distribution should become more meaningful for the auditor's decision-making. The auditor can decrease the sample size by applying the Bayesian technique at a given level of confidence, and he thus gains the maximum value from the audit sample. Ultimately, the auditor can rationally combine the subjective evidence with the quantitative sampling information, and is able to state explicitly the level of assurance for the particular proposition through the quantification of this level using Bayesian analysis.

---

28Tracy, p. 98.
Conjugate Prior Distributions

In an audit case in general, the error rate for an attribute could take on any value between zero and one, and a continuous prior distribution is thought to be more realistic than a discrete prior distribution although the process involved is a Bernoulli process. When the probability in a continuous case is computed, potential difficulties in computation are expected. To ease the computational burden, the concept of conjugate prior distributions has been developed by statisticians. A conjugate family of priors can be determined for any particular population being sampled, with the underlying process - such as the Bernoulli process - and the uniquely determined likelihood function.

A conjugate prior distribution is "conjugate" only with respect to a given likelihood function such as a binomial distribution. When sampling from a stationary and independent Bernoulli process, the conjugate family is the family of beta distribution. Felix and Grimlund used the family of normal distribution as the conjugate priors for the normally distributed population with known variances.  

---

29 Winkler and Hayes, 1975, p. 497.

The density function of the beta distribution of the error rate \( p \) with parameters \( r \) (occurrences) and \( n \) (trials), where \( n > r > 0 \), is

\[
g(p) = \frac{(n-1)!}{(r-1)!(n-r-1)!} \ p^{r-1}(1-p)^{n-r-1}
\]

The shape of the beta distribution depends on the values of \( r \) and \( n \). A useful fact is that cumulative beta probabilities and values of the beta density function can be determined from tables of the binomial distribution if \( r \) and \( n \) are integers as is true in audit sampling. \(^{31}\) For use in this research, if the prior distribution has beta parameters \( r' \) and \( n' \) and the sample consists of \( r \) error items in \( n \) items sampled, then the posterior distribution has beta parameters \( r'+r \) and \( n'+n \). \(^{32}\) The mean and the variance of a beta distribution with parameters \( r \) and \( n \) are

\[
E(p|r,n) = \frac{r}{n}
\]

and

\[
Var(p|r,n) = \frac{r(n-r)}{n^2(n+1)}
\]

\(^{31}\) Winkler and Hays, 1975, p. 499.

\(^{32}\) For a mathematical proof of this result, see Winkler and Hays, 1975, pp. 501-2.
For the posterior distribution, we can replace $r$ and $n$ with $r''$ and $n''$ respectively and can still use the same density function, mean, and variance, where

$$r'' = r' + r$$

and

$$n'' = n' + n$$

The posterior distribution with $r''$ and $n''$ as the parameters is the basis of the auditor's decision making when the Bayesian approach is used.
CHAPTER IV

THE ASSESSMENT OF PRIOR PROBABILITY DISTRIBUTIONS

Information on the relative frequency of the occurrence of an event affects the assignment of probability based on the statisticians concept that probability is a rational measure of belief. If a person who assigns subjective probabilities adopts certain principles to ensure consistent assignments, it is possible to prove that this probability measure satisfies the requirements of probability set function which has been defined by statisticians as follows:

If $P(C)$ is defined for a type of subset of the space $\Omega$, and if
(a) $P(C) > 0$,
(b) $P(C_1 \cup C_2 \cup C_3 \cup \ldots) = P(C_1) + P(C_2) + \ldots$
where the sets $C_i, i = 1, 2, 3, \ldots$, are such that no two have a point in common,
(c) $P(\Omega) = 1$,
then $P(C)$ is called the probability set function of the outcome of the random experiment.\(^{33}\)

Subjective Probability Assessment Methods

The problem a researcher in prior probability

assessment methods encounters first is the abundance of literature on the subject. Numerous techniques in the assessment of subjective probability have been employed by scholars in psychology, mathematics, statistics and related fields. As long as the technique for the quantification of judgments satisfies the probability axioms, any technique could possibly be tried for an experiment. However, there is no systematic theory about the psychology of uncertainty\(^{34}\), and the literature has not been integrated with the theory from other aspects of psychology.\(^{35}\)

Those methods which have attracted the greatest attention from scholars in this area are the use of probability and cumulative density functions, direct judgmental curve fitting, the smoothing of historical data, the application of psychometric ranking, and the use of hypothetical sample information, that is, hypothetical future samples and equivalent prior sample methods. Following is a discussion of each of these methods, and some other methods which have also been suggested. Each method is then evaluated with respect to their usefulness to the auditor.


Probability Density Function (PDF)

In the PDF method, the points on the probability decision function are assessed by direct questioning as to relative density and relative areas from which the PDF can be constructed. Two specific methods have been suggested by Schlaifer under some relative assumptions on whether the decision-maker has any quantitative evidence available to him. The two methods, judgmental curve fitting and smoothing of historical data, are discussed in turn below.

Judgmental curve fitting. The application of this method of assessing subjective probabilities requires that the following conditions be met:

(a) the probabilities of the individual values of the uncertain quantity should fall off smoothly to either side of a single mode;
(b) most of the probability should be concentrated within a fairly small interval around the mode;
(c) there should be some probability fairly far out in the tails.

If the decision-maker believes that the above conditions are met, then his judgments as to the occurrence

---

of all the possible values of the variable, for example, an error rate in the preparation of sales invoices in a company, could be represented by a mass function of the general shape as shown below.

FIGURE 1
A TYPICAL PROBABILITY DISTRIBUTION CURVE

![Probability distribution curve](image)

Smoothing of historical data. In the assessment of subjective probabilities, the assessor must use all the available knowledge at the time of the assessment task. He should use the historical data which are relevant to the assessment being made if they are available to him. The historical relative frequencies for every possible value of the unknown are plotted and a smooth curve whose general shape agrees with the decision-maker's general beliefs regarding the long-run behavior of the process, but which remains as close to the individual points as possible, is drawn. The general shape will be unimodal
and smooth if the decision-maker judges the values to be generated by a random process consisting of a basic and a random element that is the effect of a large number of small independent factors. The decision-maker then uses this estimated long-run frequency distribution as a guide for his current decisions. 37

Cumulative Density Function (CDF)

For the direct fractile assessment of the CDF, Morrison suggested the following questions to obtain the fractiles (F). 38

Question 1. At what value of the variable, F(50), do you feel that there is a 50 per cent chance that the true value of the variable will be below F(50)? (To establish the value at which CDF = 0.5.)

Question 2. Given that the true value of the variable is below F(50) at what value of the variable F(25) do you feel there is a 50 per cent


chance that the true value of the variable will be below this value? (To establish the value for CDF = 0.25.)

Question 3. Given that the true value is above $F(50)$ at what value of the variable $F(75)$ do you feel there is a 50 per cent chance that the value of the variable will be below this value? (To establish the value for CDF = 0.75.)

From the set of discrete points obtained from these, and some additional questions on the same lines, an approximate curve for the CDF is drawn. Schlaifer's two methods of assessing PDF discussed above can also be used for the CDF assessment.

Judgmental curve fitting. Very small probabilities should be assessed for the definition of the complete curve when a mass function of the typical shape is used as a representative distribution function of the decision-maker's judgments. To alleviate this difficulty to some extent, Raiffa suggested the CDF method for assessing probabilities as the method which has more intuitive appeal.\(^{39}\)

---

\(^{39}\)Raiffa, H., Decision Analysis, Introductory Lectures on Choices under Uncertainty (Addison-Wesley, 1968).
Smoothing of historical data. Smoothing of historical data can be used in either case where the probability mass or cumulative function is assessed. The problem of choosing PDF or CDF can be dealt with by analyzing the quality of the data available. Whenever small probabilities should be assessed from insufficient data, CDF will be favored by the decision-maker.

**Psychometric Ranking**

The use of a psychometric ranking technique for the assessment of a decision-maker's beliefs about business problems was suggested by Smith.\(^4\) The example used by Smith is illustrated here. The case used concerns the assessment of the expected market share for a new product by a manager. Suppose the manager breaks up the possible range of the market share into 10 intervals, say 0-10%, 10-20%, etc. Smith maintains that, rather than trying to assess the relative probabilities of each interval directly, the manager should be asked to rank the various intervals in ascending order according to their expected relative probabilities of occurrence. According to Smith, it is feasible to require the assessor to repeat

the procedure, this time using the first differences of these intervals. To translate the information imputed by the manager's rankings into relative probabilities, Smith applied the technique for the quantification of rankings designed by Kendall.\(^{41}\)

Smith suggested the use of ranking techniques for the assessment of probabilities claiming that such a method is logical, definitive and consistent. Further work in this area has come from Kruskal\(^{42}\) and Shepard.\(^{43}\) Although the application of ranking techniques may have a potential use in the assessment problem, this method was criticized by Hampton, et. al. as psychologically and intuitively meaningless to the decision-maker.\(^{44}\)

**Hypothetical Future Samples (HFS)**

The HFS method examines the effect of additional knowledge of the distribution after the subject has made

\(^{41}\)Kendall, M. G., "Ranks and Measures," *Biometrika*, 49, pp. 133-137.


\(^{44}\)Hampton, et. al, p. 27.
a single assessment for the proportion being estimated. The subject is asked to imagine what effect the additional knowledge of a given random sample would have on his original probability. If the subject has estimated the error rate in the preparation of sales invoices of a company with certain internal control strengths and weaknesses, then the imaginary sampling results could be given to the subject so that he may revise the original error rate estimate. For example, he may be asked to adjust his original estimate, say 5 per cent, to another estimate based on hypothetical sample results, like no error in a random sample of 20 sales invoices, two errors in a random sample of the same number, and so forth.

This method is based on the device of imaginary results suggested by Good. According to Good, it is by no means necessary to consider only the results of actual experiments when probabilities are estimated. In fact, Good says, it might be a mistake to do so, since the invariance theories are supposed to apply when there


46 This invariance theories of initial probability distributions for general problems were proposed by Jeffreys and Perks. See Jeffreys, H., Theory of Probability, Clarendon Press, Oxford, Third ed., 1961. Jeffreys' idea was to attribute a total probability to a region of parameter space by a rule that would lead to the same probability for the same region of parameter space when the parameters undergo a change of coordinates.
are no additional facts to obscure the issue, as there usually are in any real problem.  

Equivalent Prior Sample (EPS)

The subject expresses his prior probability in the form of equivalent prior samples. He is asked to assign values for r and n which would be equivalent to obtaining a sample in which r invoices in error of n invoices chosen at random will be observed.

Ratio Estimation

The estimation of the ratio of two probabilities has been suggested as a method which helps subjects to express extreme probabilities mainly because of the apparent reluctance on their part to make expressions in the extreme regions if they are working with bounded interval techniques.  

Direct Specification of Parameters

The direct specification of the parameters of the distribution, say means, medians and variances, may be done

47Good, 1950, p. 45.

by subjects provided that an extensive training can be
given to the subjects.\(^{49}\) The difficulty of estimating
variances and correlations, however, has been reported.\(^{50}\)

**Credibility Forecast Diagram (Potential Surprise Concept)**

Allen proposed a method of reasoning through which
a credibility region can be constructed for the estimated
probabilities,\(^{51}\) based on Shackle's credibility and
potential surprise concepts.\(^{52}\) In this context, the
criteria of project selection should be credibilities.
These credibilities do not necessarily add to one and
hence can be calculated on the available basis of incomplete
experience. A credibility forecast diagram is derived for
each variable of interest by means of the following
reasoning:

\(^{49}\)Peterson, C. R. and Beach, L. R., "Man as an
Intuitive Statistician," *Psychological Bulletin* (1967),
pp. 31-34.

\(^{50}\)Kahneman and Tversky, 1972, p. 444, and Erlick,
D. E., "Human Estimates of Statistical Relatedness,"

\(^{51}\)Allen, D. H., "Credibility Forecasts and their
Application to the Economic Assessment of Novel Research,"

\(^{52}\)Schackle, G. L. S., *Decision Order and Time in
On the basis of the information available to me at the present time I consider any value of \( x \) between \( B \) and \( C \) to be completely credible. I consider it utterly incredible for \( X \) to have a value less than \( A \) or greater than \( D \) . . . "\(^{53}\)

This is of a highly qualitative nature and the meaning is ambiguous. The actual usefulness of this approach has been criticized by psychologists.

**Range Betting Method**

This assessment technique was designed by Toda,\(^{54}\) and relies on a payoff function (or scoring rule). The assessor is asked to bet on the true value of \( x \) in the form of a range \([c, d]\), \( 0 < c < d < 1 \). The price for choosing a range of length \( r \) is \( r \). The assessor will receive the prize \( k \) if the true value of \( x \) falls within the stated range. The payoff function \( f \) is defined as

\[
f(c, d) = \begin{cases} 
  k - (d-c) & \text{if } x \in [c, d] \\
  - (d-c) & \text{if } x \notin [c, d]
\end{cases}
\]

It is assumed that the assessor wants to maximize his subjective expected payoff, i.e.

\(^{53}\) Hampton et. al., p. 29.

\[
\max_k \int_c^d g(p) \, dp + c - d.
\]
(The distribution is presumed to be absolutely continuous and unimodal.)

If the distribution is assumed to be of a certain type, say the beta, then one value of \(k\) is enough to determine the distribution parameters. If the distribution is not of a particular type, then we can let the assessor state ranges for different values of \(k\), thereby obtaining a number of ordinates of the density function.

**Empirical Tests of Various Methods**

Of the methods described in the previous section, some of the methods are relevant only to specific situations, which means that the methods are not generally applicable to other cases. A substantial portion of the experiments conducted used naive subjects, mostly psychology students taking courses or participating in laboratory experiments, and the results of these experiments cannot be applied to auditing in which only experienced or well informed auditors make decisions or estimate risk levels. In general, a great deal of this research cannot be used directly without first making tests in relation to specific questions of interest, such as error rate estimation in attribute sampling. The area is not sufficiently developed
to have a general theory which would permit direct application to accounting situations.\textsuperscript{55}

In practice the probability assessor is commonly not a statistician and is unfamiliar with the formal notions of probability. Thus, the methods used to obtain his subjective probability assessments have to be comprehensible to him.\textsuperscript{56} This requires the more elegant theoretical assessment methods such as judgmental curve fittings be abandoned in favor of a number of more basic approaches. The evaluation of various methods and experiments in practical contexts follows.

\textbf{Winkler's Study}

In practical contexts, the main contribution to the prior probability assessment task has been made by Winkler.\textsuperscript{57} Winkler selected four main assessment techniques for the measurement of an unknown proportion. The four methods were as follows:

\begin{itemize}
  \item \textsuperscript{56}Hampton et. al., 1973, p. 27.
  \item \textsuperscript{57}Winkler, 1967.
\end{itemize}
1. CDF - the fractiles of the distribution are assessed and the CDF is graphed. Morrison's methodology was followed.

2. HFS - the subjects were asked to imagine what effect the additional knowledge of a given random sample would have on his original probability.

3. EPS - each subject was asked to determine two numbers r and n such that the subject's knowledge would be roughly equivalent to having observed exactly r males in a random sample of n students in the University of Chicago student body.

4. PDF - the points on the probability density function are assessed by direct interrogation regarding relative density and relative areas from which the PDF is graphed.

According to Winkler, varying degrees of success were obtained from the use of direct (CDF and PDF) methods on the one hand and indirect (HFS and EPS) methods on the other. A feedback session was used to reduce these inconsistencies. Each subject was presented with his assessments and, after a discussion, was asked to write down one best assessment. The mere realization of the existing differences resulted in their reduction. Winkler concluded that these differences between the distributions can be explained in terms of conservatism in the subjects.
They appeared to have difficulty in making full use of all the information they had available to reduce their uncertainty, and so their assessments obtained by the direct methods were diffuse distributions.

By the same interpretation, the conservatism caused the subjects to give HFS less weight than they should, thus implying a tighter distribution than they should. Also, they gave larger sample sizes for EPS, not realizing the implied tightness of the distributions obtained. The subjects ranked the indirect methods higher than the direct methods on a clarity scale, which aims to reflect their understanding of the technique and how easy it was to apply.

Winkler's study involved the assessment of a proportion, male student proportion of the entire population rather than absolute quantity such as the thickness of steel plate. Because of this nature, his research methodology was used in two later studies of proportion (error rate) estimate which is relevant to the attribute sampling problem in auditing. As the primary purpose of this study is to find the most appropriate method of obtaining a prior probability distribution for audit decisions, the subsequently published studies on audit applications of various methods will be discussed in detail.
Corless's Study

Corless conducted an experiment in which auditors, who were active in audit work and had at least three years' auditing experience, participated in assessing prior probability distributions. A hypothetical hospital audit case was used for the subjects to answer two sets of questions on the error rate of the payroll processing. By answering the first set of questions, the subjects specified the values of the first quartile, median, and third quartile of a probability distribution. Corless constructed from this information a beta distribution and computed the values of its mean and variance. This method is similar to Winkler's CDF method.

In the second set of questions the subjects were asked to specify the probability that the actual error rate was contained in each interval. From this information, a discrete distribution was constructed and the mean, first quartile, median, and third quartile values were imputed. To determine the degree of reflection of the prior distributions of the auditors' beliefs, Corless measured the consistency between the beta and the discrete distributions by computing the discrepancy between the corresponding quartile values of the above distributions.

58 Corless, 1972, pp. 556-566.
The second method may be classified as Winkler's PDF method.

The results of Corless's experiment were as follows:

(1) Although all the subject auditors specified all information requested, there was considerable variability among the prior distributions assessed by different auditors. This may be due to the fact that the prior distributions poorly reflect the beliefs of the auditors, or due to the use of artificial audit cases in the study in which auditors were not certain as to the meaning of some of the facts given.

(2) It is possible that both the beta and discrete distributions of some auditors reasonably reflected their beliefs based on the fact that there were no discrepancies between the quartile values of these auditors.

(3) Considerable discrepancies were also found between the quartile values of the beta and the discrete distributions of some auditors. This indicates that at least one (or both) of the prior distributions does not accurately reflect the auditors' beliefs.

(4) No significant relationships were found between the measures of consistency and the differences
in the auditing experience and the statistical backgrounds of the subject auditors. This may be due in part to the assessment method which produced considerable discrepancies, and as a result, the effect of the difference in audit experience might not have been apparent.

(5) When the internal control was stronger, the prior distributions tended to have more probability concentrated on smaller amounts of error. Also, the auditors whose clients had stronger internal control tended to assess prior distributions with more probability concentrated on smaller amounts of error than did auditors whose clients had weaker internal control.

(6) Bayesian-revised distributions in general had smaller characteristic values than did the judgmentally-revised distributions.

Felix's Study

Felix conducted an experiment using Winkler's EPS method which he described as a method particularly congruent with the auditor's environment. He gave the following reason to support this argument:

59 Winkler [1967] reported that the inconsistencies tended to be smaller only for the persons with an extensive statistics background, such as professors in statistics.
...viewing qualitative evidence as equivalent to a prior sample may be relatively easy and understandable for auditors who have some experience with classical sampling. Experience in obtaining sample evidence in the form of error rates from accounting populations should make it easier to view qualitative evidence as equivalent to error rates in equivalent prior samples.60

Felix compared the results of the two assessment tasks: one using the EPS method and the other employing the quartiles assessment method which Corless had used. The results indicated that the percentage differences were a little smaller in Felix's experiment as compared to Corless's results.

While Felix's efforts in an experiment using an alternative assessment method should be commended, there are some weaknesses, both in his argument for the use of the EPS method which, he said, "seems promising" for the auditors, and in the implementation of the experiment.

Understandability of EPS. Winkler took into consideration two things when he was testing the clarity of the four techniques as perceived by the subjects: how clear and understandable the technique was to them, and how easy it was to answer the questions whether or not

60Felix, 1976, p. 802.
they liked to think in terms of the technique. Mean ratings of the techniques on the clarity scale made during each of the first and second sessions were:

<table>
<thead>
<tr>
<th></th>
<th>First Session</th>
<th>Second Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDF</td>
<td>3.68</td>
<td>2.95</td>
</tr>
<tr>
<td>HFS</td>
<td>2.00</td>
<td>1.97</td>
</tr>
<tr>
<td>EPS</td>
<td>2.79</td>
<td>2.03</td>
</tr>
<tr>
<td>PDF</td>
<td>3.05</td>
<td>2.58</td>
</tr>
<tr>
<td>All Techniques</td>
<td>2.88</td>
<td>2.38</td>
</tr>
</tbody>
</table>

(A lower number indicates greater clarity.)

In both sessions, the subjects indicated that the HFS method was more clear to them than the EPS method. The method, which may at first seem clear and understandable to the subjects, might not always to be method which is easy to use in answering the questions. The EPS method, which is used to elicit people's expectations of occurrences in a sample size n, may look in its very nature equivalent to viewing qualitative evidence in any environment where the task is to estimate the proportion of certain characteristic values (Bernoulli process).

That is, this characteristic is not a monopolistic attribute of an audit environment. Felix's experimental

\[61\text{Winkler, 1967, p. 788.}\]

\[62\text{Winkler did not indicate in his paper, however, whether these differences between the methods were statistically significant or not.}\]
result is self-contradictory because he reported that "eight out of the ten subjects commented on the intuitive appeal of viewing qualitative evidence as being equivalent to a prior sample. Unfortunately, this appeal was not so apparent when they faced the assessment task." 63

Less dispersion of distributions. The major significance of Felix's experiment was that the percentage differences between the EPS method and the direct assessment of the quartiles method were smaller than in the Corless study. However, this result is not very impressive, because it had already been observed in Winkler's study that the indirect techniques produced smaller credible intervals (both 50 percent and 98 percent), and hence tighter distributions, than did the direct techniques. Felix used the same EPS interview format used by Winkler and the assessment was of the proportions of the characteristic in question in both studies, although the case which was used in each experiment was different. Also, "some" difficulty in using the direct fractiles method was experienced by all of the auditors in Felix's experiment and one auditor (10 percent of Felix's sample) thought the first and third quartiles as the location of a 50 percent credible interval centered on the median.

Accordingly, it is not very difficult to understand that an easier method (EPS) produced less inconsistent assessments.

**Implementation of the experiment.** One of the major difficulties in implementing an experiment of this nature is to have the subject auditors for an adequate time period and to motivate and maintain their interest. However, a sample size of ten is too small to draw a significant conclusion from the results of the experiment. Also, the personal background factors, such as the level of training in statistics and the amount of audit experience, were not considered by Felix. He gave a reason for this deficiency by quoting Winkler's statement that education and training in statistics can be a significant factor only when some subjects have "extensive" backgrounds while others do not. However, he did not consider the possible impact of the differences in the amount of auditing experience on the assessments. He did not disclose the amount of the actual auditing experience of the subject auditors. He simply mentioned "ten practicing auditors." In the Corless study, the impact of audit experience on the resulting distributions was found to be insignificant when auditors with at least three years experience were used as subjects.

---


65 Corless, 1972, p. 562.
However, this result cannot be generalized to the case in which the assessments were made using an easier, more understandable indirect method (EPS or HFS). In other words, if the difficulty in understanding the method used by the auditors becomes less significant, the effect of the other factor, the amount of the audit experience, which might have been mixed with the effect of the relatively modest statistical background of the auditors, may become apparent. Thus there is still the possibility that, as far as the assessment of priors is concerned, the marginal improvement in assessment ability for the additional audit experience beyond three years may not be as great as that for the increased experience below three years.

**Evaluation of "the Four Methods"**

Based on the nature of the attribute which Winkler considered in his experiment; that is, a proportion rather than absolute quantity, and the promising nature of the practical usefulness of Winkler's four methods (CDF, PDF, EPS, and HFS), three methods (CDF, PDF, and EPS) from Winkler's study have been tested in the auditing context by Corless and Felix. Because of this importance of these four methods, a more detailed discussion of them is desirable.
CDF

This is actually the most often suggested technique in the literature. As was explained earlier in this chapter, this is the technique of assessing fractiles by means of successive subdivisions. The assessor is asked to subdivide a given interval into two equally likely parts. This subdivision is continued until there are enough fractiles to indicate the shape of the distribution function.

The advantage of this technique is that it only require the assessor to make judgments about whether or not two events are equally likely. The concept of 'equally likely' ('fifty-fifty') has some appeal even to the statistical novice. In Winkler's study, it was found that his subjects were less sure of working with CDF. The use of PDF was more intuitively appealing than CDF. Hampton et al. concluded, however, that the direct fractile assessment method may well be the approach to adopt for decision analysis in a business environment.

---


where the estimate may be of a quantity rather than a proportion. In the audit environment where error rate estimation is involved, the use of CDF is not appropriate as was observed in Corless's study and Felix's experiment.

**PDF**

PDF was more intuitively appealing to the subjects in Winkler's study than CDF. However, according to Winkler, the use of this method by an assessor in a consistent manner requires more training and knowledge in probability assessment than is required in the use of two indirect methods (HFS and EPS). However, the same cannot be said about the assessments by auditors with adequate amounts of experience until tested empirically.

**EPS**

The clarity of the EPS method was rated higher than the PDF and CDF methods by Winkler's subjects but lower than the HFS method. In a similar experiment by Schaefer and Borcherding, subjects found the EPS technique more difficult than the successive subdivision technique applied to the cumulative distribution, until after

---

68 Hampton et al., p. 28.

substantial training was provided. In Felix's experiment which involved the same two techniques (EPS and a variation of CDF) Schaefer and Borcherding used and the same kind of assessment characteristic (assessment of a proportion of a certain characteristic), the subjects found EPS easier than the direct fractiles assessment technique. This may be due to the fact that viewing qualitative evidence as equivalent to a prior sample may be easy for auditors who have some experience with classical attribute sampling.

The major difference in the design of the experiment is the Schaefer/Borcherding study and Felix's study was the choice of subjects: S & B used psychology students enrolled in a course while Felix used experienced professional accountants who in their auditing process must frequently evaluate sample evidence.

**HFS**

The HFS technique, the device of imaginary results by Good, examines the effect of additional knowledge of the distribution on the initial assessment of the proportion in question. This method measures the reactions of the subject to imaginary random sample results based on the current judgments. In other words, this is a backward application of Bayes' theorem in that the technique asks, if the subject has a posterior, what is the respective prior.
The previous section of conjugate prior distributions showed that the density function of the beta distribution for the error rate \( p \) with parameters \( r \) (occurrences) and \( n \) (trials), where \( n > r > 0 \), is

\[
g(p) = \frac{(n-1)!}{(r-1)!(n-r-1)!} \frac{r^r (1-p)^{n-r-1}}{p} , \quad 0 < p < 1.
\]

In the meantime, a pair of independent hypothetical samples \( (r_1, n_1) \) and \( (r_2, n_2) \), which represent the event occurring \( r \) times in \( n \) trials, with the corresponding estimates \( p_1 \) and \( p_2 \) could be used to estimate the parameters \( a \) and \( b \) in the beta density function

\[
g_1(p) = p^{a-1}(1-p)^{b-1} , \quad 0 \leq p \leq 1, \quad a > 0, \quad b > 0,
\]

which is analogous to the above \( g(p) \) function. This works as follows. The conditional distribution for \( r \), given the probability \( p \), is an ordinary binomial distribution with density function.

\[
f_{r|p}(r|p) = \binom{n}{r} p^r (1-p)^{n-r}, \quad x = 0, 1, \ldots, n.
\]

The posterior distribution of \( p \), given \( r \), is given by Bayes' theorem as

\[
f_{p|r}(p|r) = f_p(p) f_{r|p}(r|p) = p^{a+r-1}(1-p)^{b+n-r-1}
\]
The posterior is thus another member of the beta family. The best estimate of the probability assessor when his loss function is quadratic is the posterior mean \( p'' = \frac{a+r}{a+b+n} \). He will choose other estimates with other loss functions but they will be close to \( p'' \) if \( a, b, \) and \( n \) are not too small. The values of \( a \) and \( b \) could be obtained by solving the equations

\[
p_i = \frac{a+r_i}{a+b+n_i}, \quad i = 1, 2.
\]

If we have \( k \) samples we could take them two by two and calculate the corresponding values of \( a \) and \( b \). The final estimates of \( a \) and \( b \) could be taken as the average values of the \( \binom{k}{2} \) estimates. The detailed application of this method will be explained in a practical manner in Chapters Five and Six.

Also, for the HFS method conflicting results were observed in separate studies by Winkler and Stael von Holstein concerning the clarity of the method. In Winkler's study, subjects rated the HFS method as the easiest to understand and use among PDF, CDF, EPS, and HFS in both of the two sessions. The distributions assessed by using the HFS and the EPS methods were consistently tighter than

\[70\] Stael von Holstein, 1970, p. 57.
the other two although the subjects gave hypothetical sample results less weight than they should have, presumably because of conservatism. 71

Stael von Holstein reported that his subjects had serious difficulty in interpreting those hypothetical sample results. Frequently his subjects' estimates of parameters (a and b above) were incompatible under the assumption of a beta distribution. A further analysis revealed that the devices which had the true probabilities greater than 0.5 (not like low error rate in auditing) had a much greater number of inconsistencies than the other devices. Apparently his subjects needed more sensitive hypothetical samples to reflect the sample information in their SPD assessments - probably a larger number of samples and more varied samples, as discussed by Stael von Holstein. 72

The conflicting results reported by Winkler and Stael von Holstein confirm that there is no single method which is appropriate in every situation. A careful screening and test of methods is necessary to find the most appropriate method in a given environment.


CHAPTER V
THE PURPOSE OF THE STUDY AND THE METHODOLOGY

The Purpose

As has been indicated, study of the appropriate SPD assessment technique(s) is essential prior to any attempt to train auditors to use the Bayesian technique in the sampling or decision theoretic approach in audit decision making.  

The primary purpose of this study is to test the SPD elicitation techniques which seem to be most congruent with audit environment and auditors' backgrounds, and to

---

73 There are two different aspects in what researchers call "goodness" of probability assessments, that is, whether an assessment is of good quality or of bad quality. First, an assessor should have some knowledge of probability theory and should be able to make consistent assessments of his beliefs according to probability theory. Second, the assessor should have practical knowledge of what he is going to evaluate. In internal control evaluation, for instance, this will be past audit experience. Expertise in probability assessment was termed as the "normative standard of goodness" and expertise in the domain in which the assessments are made was referred to as the "substantive standard of goodness" by Winkler and Murphy. See R. L. Winkler and A. H. Murphy, "'Good' Probability Assessors," Journal of Applied Meteorology, 7(1968), p. 753. The goodness criterion of primary concern in this study is the normative standard of goodness.
investigate the basic qualifications which are necessary for auditors to be able to assess their prior distributions in a reasonably consistent manner.

More specifically, three purposes exist with respect to the primary purpose:

(1) To investigate the auditor's perception and acceptance of hypothetical samples techniques (EPS and HFS) of SPD elicitation. Felix reported the intuitively appealing nature of the EPS method to the subject-auditors based on his experiment. Felix's study, however, was conducted "to provide some evidence on the feasibility of using the EPS method." The study did not demonstrate the relative advantage of using the EPS method to produce more consistent assessments by auditors. The study merely confirmed a priori promise of the EPS method that "experience in obtaining sample evidence in the form of error rates from accounting populations should make it easier to view qualitative evidence as equivalent to error rates in equivalent prior samples."74

The Felix study indicated that an intuitively appealing method (EPS) was feasible with limited training. As a result of this training, auditors' assessments of priors were "somewhat" more consistent than Corless's

---

74 Felix, 1976, p. 802.
results which were obtained without giving any training to the subject-auditors prior to assessments.

Like the EPS method, the HFS technique also has both a priori and empirical support as a technique worthy of testing in an auditing environment. The auditor with an adequate audit background usually has sufficient experience in evaluating sampling results. The HFS technique may help the auditor refine his prior probability distribution as to the adequacy of internal control, and past and current results of various audit tests by making the auditor ask himself questions such as: "Suppose I had taken a sample of 100 items from this account and exactly 10 items were found to be in error. Now what is the probability that one additional item to be sampled at random from the entire population is in error?" The auditor could check the sensitivity of his judgment using this method, potentially resulting in a more consistent assessment.

Any experiment testing the auditor's perception and acceptance of hypothetical samples methods should use an adequate sample size. This is essential to the drawing of meaningful conclusions as to the appealing nature of methods and as to any impact of auditors' backgrounds on the perception of methods and the consistency of assessed distributions. Felix used a sample size of only ten and
this limited sample size made the analysis of the above impacts not feasible, as Felix acknowledged. 75

(2) To investigate the consistency in SPD assessments resulting from the use of two hypothetical samples techniques. The primary thrust of Felix's findings is that, in comparison with Corless's results, average differences of assessed quartile values as percentages of the beta quartiles are "somewhat smaller." Felix attributed the increased consistency between methods to the limited training the subjects received in his study.

It is not possible, however, to determine the extent of the increased consistency in assessments attributable to the use of an appealing method and the extent attributable to the training which the auditors received in Felix's study. From the previous literature review it is almost certain that training will increase consistency in assessed SPDs with the use of the same assessment techniques. Accordingly, it is highly probable that the increased consistency might be entirely attributable to the training and not to the method used.

In this study, the consistency will be measured in two aspects; discrepancies between the EPS method and the

---

75 Refer to the letter from Felix in this respect which is appended later in Appendix 6.
discrete PDF method, and between the HFS method and the
discrete PDF method. The reason for choosing two hypothe­
tical samples techniques (EPS and HFS) has been explained
in detail in the previous sections. For a multifactor
study, the discrete PDF method, which was used in both
the Winkler and Corless studies, has been selected because
(a) the PDF was rated to be more understandable by the
subjects in Winkler's study as discussed previously, and
(b) to facilitate a more meaningful comparison with
Corless's results. More difficult techniques would
generally produce more inconsistent assessments than easier
techniques. Therefore, differences in auditors' modest
backgrounds in probability theory and sampling and in
audit experience would not play any role in discriminating
strong background-holders (and more experienced auditors)
from weak background-holders (and less experienced auditors)
if even a subject-auditor with a relatively strong back­
ground in probability has difficulty in applying a method
(the direct fractiles method in the Corless and Felix
studies). In other words, the choice of a difficult
assessment method would obscure relative merits of stronger
backgrounds.

(3) **To investigate the effect of personal background
factors on discrepancies in assessed SPD and to find basic
qualifications necessary for reasonably consistent
assessments of SPD.** In Felix's study, only ten
subject-auditors with essentially homogeneous training and personal backgrounds were used and "there was almost no difference in the amount of auditing experience or statistical backgrounds of the subjects." Therefore, any analysis on the above relationship was impractical and not feasible.

Corless observed no significant relationship between auditors' backgrounds and discrepancy in assessed SPDs. It is likely that the methods used were not fully understood by the subjects or very difficult to apply (especially direct fractiles method), which overwhelmed the slight differences in their backgrounds. All the auditors who participated in Corless's study had three or more years of auditing experience and this may have prevented him from finding any relationship, because three years of audit experience may be above the "threshold."

As to the normative standard of goodness which was discussed in the first part of this chapter, there certainly exist the minimum qualifications required for reasonably consistent (or reliable) assessments of SPD. Thus, a study of the basic knowledge requirements in probability, sampling and/or audit experience is necessary before the Bayesian approach can be used in a practical context.

76 Letter from Felix, Appendix 6, p. 1.
The Methodology

Case and questionnaire. A questionnaire containing two audit cases and questions on the error rates in sales invoice preparation was administered to 36 auditors who have more than one year of actual audit experience.

The first audit case described in some detail the organization, history, management, and operations of a hardware wholesaler of small size. Internal control procedures for sales and collection, and the prior year's (1976) audit procedures for the billing function were described together with audit results. Those results included statistical sampling results for the attribute of the accuracy of sales invoice preparation. For the year being studied (1977), some improvements in internal accounting control for the billing function were made, but some weaknesses remained. Unaudited 1977 key financial figures were provided to give the subjects an idea as to the size of the company and business volume.

A medium-size hardware wholesaler case which was developed from the first audit situation was used as the second case rather than using two attributes for the same company billing function (Felix's study). This was done to reflect a possible impact of client size on the assessed SPD for error rate. In overall size, the second company was ten times as large as the first company in
all categories. A stronger internal verification procedure was assumed as compared to the first company, but there were still some control weaknesses.

For the EPS method, Winkler's EPS questionnaire was used with necessary modifications in the attribute being evaluated, characteristics of the case, and sampling implications. When the experimenter was testing the initial questionnaire, the reviewers commented on the possibility of the subjects viewing the denominator n of the equivalent prior sample r/n as the final audit sample size. A caution as to this was included in the questionnaire. Also, there was the danger that subject-auditors might use the previous year's sample size as the given value of the denominator n and simply imagine the magnitude of the numerator r on the basis of the given value n. A clear description in this respect was also made.

When the hypothetical samples were designed for the HFS method, the initial random sample sizes were increased to larger sample sizes because some reviewers commented that those hypothetical sample sizes were too small to let the subject-auditors revise their initial estimate of error rates. Later survey confirmed this possibility. A significant variation in hypothetical sample results was necessary to stimulate the subjects' imagination of possible sample results.
Corless's discrete PDF technique was used with a change in the intervals of error rates estimate - more intervals were made to enhance the accuracy of imputing quartile values from the discrete histograms of assessed cumulative probabilities.

At the end of the questionnaire, questions on the subjects' backgrounds in probability, sampling, statistics, audit experience, continuing education in sampling, and in-firm training in statistical sampling, and client characteristics were asked. To elicit their perception and acceptance of each method, a seven-point scale was provided. (One: very easy, four: moderate, and seven: very difficult.) A seven-point scale is the maximum range scale in ordinary ordinal scale assessments frequently used in a survey such as this. Seven points rather than five or four was used to minimize possible "ties" in the score so that a better nonparametric statistical analysis would be possible for varied ranks.

Two review sessions were made with four CPAs prior to administering the questionnaires to insure understanding of the questionnaire design. These reviews proved to be useful as discussed above. A copy of the questionnaire is presented in Appendix Four.

Subjects. CPAs in Baton Rouge, Louisiana and Buffalo, New York participated in the survey through their respec-
tive offices. Questionnaires were sent to managing partners of "big Eight" firms which were willing to participate in the experiment. Partners administered the questionnaires to staff auditors who had more than one year of actual audit experience. Each respondent then initialed the provided envelope and returned the questionnaire to the experimenter individually. When the questionnaires were provided, each firm's name had been marked on the envelopes, and thus later follow-up was possible on those non-responses. Thirty-two out of thirty-seven questionnaires were returned and four responses which were obtained initially were added to the sample results making the sample size thirty-six. Four initial responses were included, because (1) the four responses were obtained based on the revised questionnaire, not the first draft, (2) only minor revisions were made in the revised questionnaire, (3) many of the thirty-two respondents also discussed the case with the author (in person or on the phone) before they completed the questionnaire, and (4) the results obtained from the four respondents were not much different from those of other CPAs with comparable backgrounds. Especially, there was no substantial increase in consistency in assessments as discuss in Chapter Six.

77 Ernst & Ernst, Deloitte Haskins & Sells, Peat, Marwick, Mitchell, and Price Waterhouse were those "big Eight" firms.
Methodology for purpose (1). Mean comparisons were made for the subjects' ratings of each assessment method with regard to easy-to-understand and easy-to-apply categories. In addition, the Friedman test for multiple comparisons was made to investigate the appeal of the hypothetical samples (EPS and HFS) methods to auditors. The Friedman test is concerned with the homogeneity of ranked objects rather than independence of ranks. The null hypothesis in this case is that the average ratings among 36 subjects for three methods do not have any significant difference. That is

\[ H_0: \ u_{EPS} = u_{HFS} = u_{PDF} \]

Where \( u_{EPS} \) indicates the average rating (easy-to-understand or easy-to-apply) for the EPS method. The Friedman test allows an analyst (a) to check whether there is any significant difference among different categories (methods), and (b) to find out which category differs significantly from which others.

To investigate whether the overall perceptions of the auditor group are significantly different from a non-auditor group with respect to the three methods being tested, 27 accounting majors who were enrolled in the senior auditing course (Fall 1978) at State University of New York at Buffalo were asked to assess SPDs using the
three methods and to rate each method on the same seven-point ordinal scale. A comparison between the auditor group and the student group was made for the subjects' perceptions of each method being studied. The results are analyzed in Chapter Six.

Methodology for purpose (2). Consistency in SPD assessment was measured as the amount of discrepancies (for inconsistency, accurately) between quartile values of the beta distribution obtained by the EPS method and the corresponding quartile values of the discrete PDF method. Those discrepancies were obtained by imputing the cumulative probability densities of the histogram drawn for the intervals of various error rates. The same was done for comparison of the HFS method and the discrete PDF method.

For both cases A and B, average discrepancies were obtained at each quartile; first quartile, median, and third quartile. To measure relative discrepancies for each case, average discrepancies were divided by respective quartile values. To illustrate the range of the difference, the greatest discrepancies were presented.

To check the amount of agreement among the subjects, differences in locations of median ($F_{.50}$) responses were presented, while 50 percent credible intervals were used to investigate the relative dispersion of the SPDs for the three techniques with a given subject
and a given attribute. As Winkler reported, a measurement problem is encountered in this type of study. If true SPDs existed, experimenters could determine measures of "good fit" to the true SPDs. To compensate, the above measures of comparison have been suggested.  

Though the same kind of assessments were obtained from the student group as to error rates, the comparison made for purpose (1) was not attempted because student error rate estimates do not reflect the assessors' prior beliefs as they do not have any actual audit experience.

**Methodology for purpose (3).** Two multiple regressions were run with $Y_1$ and $Y_2$ as dependent variables in two equations and $X_1$, $X_2$, $X_3$, and $X_4$ as explanatory variables.

$Y_1$ $=$ Discrepancy between medians of beta and discrete distributions estimated based on the assessments using the EPS method and the discrete PDF method respectively. That is, 

$$Y_1 = \text{EPS}_F^{.50} - \text{PDF}_F^{.50}.$$

$$Y_2 = \text{HFS}_F^{.50} - \text{PDF}_F^{.50}. \quad (\text{Both } Y_1 \text{ and } Y_2 \text{ are for the case A})$$

$X_1$ $=$ Subjects' auditing experience in number of years.

---

78Winkler, 1967, p. 786.
\( X_2 = \) Subjects' statistical background expressed in number of courses they have taken.

\( X_3 = \) Subjects' sampling knowledge expressed in number of hours of instruction in statistical sampling they have received from their current and previous affiliations.

\( X_4 = \) Subjects' evaluation of the size (sales) of the majority of their clients as compared to the company A and the company B in the case.

In addition to these four background factors, information on continuing education in sampling and the strength of the internal control of the majority of their clients as compared to the control of two companies in the case was obtained through the questionnaire. The results of these two were almost homogeneous among all the subjects and these two factors were dropped from the regression analysis. To investigate the impacts of each individual background factors on the inconsistency (discrepancy in the equation) in assessed SPDs, four separate, simple bivariate regressions were run on two dependent variables, \( Y_1 \) and \( Y_2 \).

As a result we have the following regressions:

(a) \( Y_1 = a + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + e_1 \)

(b) \( Y_2 = a' + b'_1X_1 + b'_2X_2 + b'_3X_3 + b'_4X_4 + e_2 \)
(c) \[ Y_1 = A_1 + B_1 X_1 + E_{11} \]
\[ Y_1 = A_2 + B_2 X_2 + E_{12} \]
\[ Y_1 = A_3 + B_3 X_3 + E_{13} \]
\[ Y_1 = A_4 + B_4 X_4 + E_{14} \]

(d) \[ Y_2 = A'_1 + B'_1 X_1 + E_{21} \]
\[ Y_2 = A'_2 + B'_2 X_2 + E_{22} \]
\[ Y_2 = A'_3 + B'_3 X_3 + E_{23} \]
\[ Y_2 = A'_4 + B'_4 X_4 + E_{24} \]

where: \( a, a', A_i, A'_i \) are regression intercepts (\( i=1,...,4 \))

\( b_i, b'_i, B_i, B'_i \) are regression slopes (\( i=1,...,4 \))

\( e_1, e_2, E_{ij} \) are random disturbance terms

\((i=1,2 \text{ and } j=1,...,4)\)

Among the discrepancies between three quartiles of the beta and discrete distributions, the discrepancies between median values \( (F_{.50}) \) were selected to determine the inconsistency in assessed SPDs for the following reasons. One, Winkler's study indicated that median values \( (F_{.50}) \) represented the subjects' inconsistencies best - the most consistent assessor had almost no discrepancy at the median values of four SPDs on the basis of the four different methods while there were still discrepancies at the first and third quartile values. Two, Winkler also reported that the most distinguishable feature in the graphed curves of SPDs of consistent and inconsistent assessors was whether all the curves had the same median
values. This fact was reconfirmed when the SPDs were graphed in the present study to impute quartile values. Three, if point estimates rather than probability distributions were made for each attribute being studied, then the estimates should represent the median values rather than two other quartiles. As long as each assessor obtained desired median values, it seems that they cared less about two other quartile values.  

Summary. In summary, the specific purposes and methodologies are as follows:

<table>
<thead>
<tr>
<th>Specific purposes</th>
<th>Methodologies</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Investigate the auditor's perception and acceptance of hypothetical samples techniques.</td>
<td>(1) Mean comparisons of ratings.</td>
</tr>
<tr>
<td></td>
<td>(2) Friedman test.</td>
</tr>
<tr>
<td></td>
<td>(3) Contrast the auditor group with the student group.</td>
</tr>
<tr>
<td>2. Investigate the consistency in SPD assessments for hypothetical samples techniques.</td>
<td>(1) Discrepancy table for quartiles.</td>
</tr>
<tr>
<td></td>
<td>(2) Locations of median responses.</td>
</tr>
<tr>
<td></td>
<td>(3) 50 Percent credible intervals.</td>
</tr>
<tr>
<td>3. Investigate the effect of personal background factors on discrepancies.</td>
<td>(1) Multiple regressions.</td>
</tr>
<tr>
<td></td>
<td>(2) Simple regressions.</td>
</tr>
</tbody>
</table>

CHAPTER VI
ANALYSIS OF THE RESULTS

The Auditor's Perception of Hypothetical Samples Techniques

Mean ratings. After the subject-auditors assessed distributions, they evaluated the appealing nature of the two hypothetical samples methods and the discrete PDF method by assigning one of the seven ordinal scale values. The results are summarized in Table 1.

The two hypothetical samples techniques were perceived as easier than the direct assessment technique in both contexts; easy to understand and easy to use.

Of the two hypothetical samples techniques, the HFS method was rated as an easier technique to understand than the EPS technique. Also, in the how-easy-to-use context, the HFS method was rated lower (easier) than the EPS method, and the difference in the mean ratings was even greater as compared to the how-easy-to-understand context.

Even though experienced auditors were used as subjects, the overall ratings were a little higher in both contexts than those in Winkler's study. There may be two different explanations for this result: One, the auditor-subjects did not receive pre-assessment training in the present study and the difference might be attribu-
TABLE 1
MEAN RATINGS OF THE TECHNIQUES
ON THE CLARITY SCALE*
(THE AUDITOR GROUP)

<table>
<thead>
<tr>
<th>Technique</th>
<th>How easy to understand</th>
<th>How easy to use actually</th>
</tr>
</thead>
<tbody>
<tr>
<td>EPS</td>
<td>2.78</td>
<td>2.93</td>
</tr>
<tr>
<td>HFS</td>
<td>2.69</td>
<td>2.56</td>
</tr>
<tr>
<td>PDF</td>
<td>3.14</td>
<td>3.33</td>
</tr>
</tbody>
</table>

*A lower number indicates greater clarity.
table to better understanding of the methods by Winkler's subjects because of the training; and two, auditors may be more conservative in assigning a perception rating.

Friedman test. To test the null hypothesis that average ratings for three methods do not have any significant differences, the Friedman test was conducted. The Friedman test makes it possible for a researcher (a) to investigate whether there is any significant difference among different categories (methods in the study) and (b) to check which method differs significantly from which others.

The mean ratings of the three methods (treatment effects) are the means of the ranks of the n (three in this study) treatment effects for the data, and can be used as estimates of the ranks \( R_j \)'s of the corresponding treatment effects in the population. That is, \( R_j/k \) (k: number of rows in the k x n table) is an estimate of the rank of \( u_j \) for \( j = 1, \ldots, m \). Simultaneous multiple comparisons with an overall level of significance can be made based on the estimated treatment ranks. By taking all possible differences between rank sums of two treatments (column tables), the probability is at least \( 1-\alpha \) that the following inequality is satisfied by all pairs \( (R_i, R_j) \) for \( 1 \leq i \neq j \leq n \):\(^80\)

---

The constant $z$ is the quantile point of the normal curve that corresponds to a right-tail probability of $\alpha/n(n-1)$, since the total number of comparisons is $n(n-1)/2$. $z$ can be obtained from a normal distribution table for any $n$ and $\alpha$. At the overall level $\alpha$, all pairs of differences of column sums that are larger than the right-hand side of the above inequality expression are significantly different pairs. The direction of difference is determined by the sign of $R_i - R_j$.\(^{81}\)

The test statistic used in this test is\(^{82}\)

$$12S = 12 \sum_{j=1}^{n} R_j^2 - 3k^2 n(n+1)^2$$

and,

$$Q = \frac{12S}{kn(n+1)}.$$

The data for statistical analysis are presented in Table 2. From the above equations, the following is calculated (for easy-to-understand category).

$$12S = 12(67.5^2 + 64.0^2 + 84.5^2) - 3(36)^2 3(3+1)^2$$

$$= 2502,$$

\(^{81}\)Gibbons, 1976, p. 313.

\(^{82}\)Gibbons, 1976, p. 306.
TABLE 2

\textit{k x n TABLE OF RANKS (k=36 and n=3)}

\textit{FOR FRIEDMAN TEST}

<table>
<thead>
<tr>
<th>Subject</th>
<th>EPS</th>
<th>HFS</th>
<th>PDF</th>
<th>EPS</th>
<th>HFS</th>
<th>PDF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2.5</td>
<td>1</td>
<td>2.5</td>
</tr>
<tr>
<td>2</td>
<td>1.5</td>
<td>1.5</td>
<td>3</td>
<td>1.5</td>
<td>1.5</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>1.5</td>
<td>1.5</td>
<td>3</td>
<td>1.5</td>
<td>1.5</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>1.5</td>
<td>1.5</td>
<td>3</td>
<td>1.5</td>
<td>1.5</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1.5</td>
<td>1.5</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>2.5</td>
<td>1</td>
<td>2.5</td>
<td>2.5</td>
<td>1</td>
<td>2.5</td>
</tr>
<tr>
<td>8</td>
<td>1.5</td>
<td>1.5</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>1.5</td>
<td>1.5</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1.5</td>
<td>1.5</td>
<td>3</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>12</td>
<td>1.5</td>
<td>1.5</td>
<td>3</td>
<td>1.5</td>
<td>1.5</td>
<td>3</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>2.5</td>
<td>2.5</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>14</td>
<td>3</td>
<td>1.5</td>
<td>1.5</td>
<td>3</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>16</td>
<td>1.5</td>
<td>1.5</td>
<td>3</td>
<td>1.5</td>
<td>1.5</td>
<td>3</td>
</tr>
<tr>
<td>17</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>2.5</td>
<td>2.5</td>
<td>1</td>
<td>2.5</td>
<td>2.5</td>
</tr>
<tr>
<td>19</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>21</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>22</td>
<td>2.5</td>
<td>1</td>
<td>2.5</td>
<td>2.5</td>
<td>1</td>
<td>2.5</td>
</tr>
<tr>
<td>23</td>
<td>3</td>
<td>1.5</td>
<td>1.5</td>
<td>3</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>24</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>25</td>
<td>2.5</td>
<td>1</td>
<td>2.5</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>26</td>
<td>1.5</td>
<td>1.5</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>27</td>
<td>2.5</td>
<td>2.5</td>
<td>1</td>
<td>2.5</td>
<td>2.5</td>
<td>1</td>
</tr>
<tr>
<td>28</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>29</td>
<td>2.5</td>
<td>2.5</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>30</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>31</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>32</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>33</td>
<td>1</td>
<td>2.5</td>
<td>2.5</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>34</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>35</td>
<td>1</td>
<td>2.5</td>
<td>2.5</td>
<td>1</td>
<td>2.5</td>
<td>2.5</td>
</tr>
<tr>
<td>36</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

\[ \sum R_j: \] 67.5 64 84.5 70.5 59.0 86.5
and,

\[ Q = \frac{2502}{36(3)(4)} = 5.79, \quad \text{with df} = 3-1 = 2. \]

From the chi-square distribution table the right-tail probability is found to be between .10 and .05. Accordingly, the null hypothesis is rejected at the 10 percent level of significance and there exists significant difference in the subjects' perceptions of the three different methods.

For the multiple comparisons for the pairs of treatment effects (three methods), the following statistic is computed:

\[ |R_i - R_j| \leq 1.834 \sqrt{\frac{36(3)(4)}{6}} = 15.5. \]

For the three comparisons to be made (EPS-HFS, HFS-PDF, and PDF-EPS), and the overall significance level of .20 for the multiple comparisons, a z value of 1.834 was obtained from Fisher and Yates's Table of Critical z Values for Multiple Comparisons.\(^83\) The above statistic, 15.5, was compared with the observed difference between each pair of column sums. The difference pair of HFS-PDF was significant at the overall significance level of .20 but the EPS-PDF was not. When the overall significance

level of .30 was used, the two pairs, EPS-PDF and HFS-PDF, both had significant differences.

Values of $|R_i - R_j|$ for the easy-to-apply category were 16.0 (EPS-PDF), 11.5 (EPS-HFS), and 27.5 (HFS-PDF) as compared to the above test statistic 15.5. Therefore, in this category also, the two hypothetical samples techniques were perceived as easier than the discrete PDF method, when the overall significance level of .30 was used for the multiple comparisons analysis. However, in both contexts there was no significant difference between methods at low significance levels such as .10. Relatively high overall significance levels are used in multiple comparisons like this.

**Group comparison.** The student group was asked to assess SPDs using the three methods and to rate each method on the same seven-point clarity scale in an attempt to investigate whether the overall perceptions of the auditors group are significantly different from a non-auditor group as to the appealing nature of the hypothetical samples techniques. Mean ratings of the techniques given by the group are presented in Table 3.

There was virtually no difference among the ratings given by the subjects in the student group. The average ratings for the two hypothetical samples methods were lower than those for the PDF method in both categories.
There was, however, no statistical significance for the observed, slight differences. For the twenty-seven observations, ranks which were used in the analysis of the auditor group were assigned and the column totals \( \Sigma R_j \) were obtained. These totals were almost the same across the columns, indicating clearly no significance in the differences.

A scanning of the student backgrounds in probability and statistics revealed that all of them had the same number of courses (two courses) in statistics, which was almost identical to the average amount of statistics education the subject-auditors received. Accordingly, their use of PDF caused about the same problem as the auditors' use of this discrete, direct assessment method. The students did not really have any experience (although some of them were working part-time for public accounting firms) on which they could base their judgments, and the EPS and HFS methods were not very "appealing" to them as compared to the PDF method. Many subjects had great difficulty in determining the denominator \( n \) of \( r/n \), and it seemed that they simply used the same sample size which the previous auditor in the case had used. This result supports the initial hypothesis that the hypothetical samples techniques have clear "appeal" to auditors with actual audit experience.
### TABLE 3

**MEAN RATINGS OF THE TECHNIQUES ON THE CLARITY SCALE**

*(THE STUDENT GROUP)*

<table>
<thead>
<tr>
<th>Technique</th>
<th>How easy to understand</th>
<th>How easy to use actually</th>
</tr>
</thead>
<tbody>
<tr>
<td>EPS</td>
<td>3.69</td>
<td>4.14</td>
</tr>
<tr>
<td>HFS</td>
<td>3.84</td>
<td>4.02</td>
</tr>
<tr>
<td>PDF</td>
<td>3.92</td>
<td>4.19</td>
</tr>
</tbody>
</table>
The consistency in SPD Assessments
Discrepancies for quartile values. To investigate the
consistency in SPD assessments resulting from the use of
two hypothetical samples techniques, discrepancies between
quartile values of prior distributions assessed by using
the EPS, HFS, and discrete PDF methods were measured and
are presented in Tables 4 and 5. Quartile values of beta
prior distributions were determined in the following
manner.

Quartile values of beta prior distributions assessed
by the use of the EPS method were obtained directly from
Tables of Beta Cumulative Functions which are presented
in Schlaifer's Analysis of Decisions under Uncertainty.\textsuperscript{84}
Quartile values of beta prior distributions for the HFS
method were determined following Winkler's format. The
HFS questions produced five answers, which could be
converted to five hypothetical future sample sizes as
follows.

The two numbers, \(r'\) and \(n'\), were the parameters
to be determined and used for a beta distribution assessed

\textsuperscript{84}Schlaifer, Robert, Analysis of Decisions under
sizes of \(r'\)s and \(n'\)s in beta distributions, fractiles are
given as follows: \(F_{.001}, F_{.01}, F_{.05}, F_{.1}, F_{.25}, F_{.5},
F_{.75}, F_{.9}, F_{.95}, F_{.99}, F_{.999}\). (\(F\: \text{cumulative probability}\)
by each subject using the HFS method. First, the subject was asked to estimate the error rate in the client's sales invoice preparation based on the information given in the case. This estimated error rate ($p^*$) should be equal to the proportion $r'/n'$, because both of them reflect the subject's judgment of the same attribute. Now $r' = p*n'$. In the meantime, for each hypothetical question a revised error rate estimate was made and there were five revised estimates ($a, b, c, d, \text{ and } e$). For the first set of hypothetical sample with $r$ and $n$, then, the following relationship held: 

$$\frac{r' + r}{n' + n} = a.$$  

And, for the five hypothetical sample results, the following equations were obtained:

1. $$\frac{r'}{n' + 20} = a$$
2. $$\frac{r' + 2}{n' + 20} = b$$
3. $$\frac{r' + 1}{n' + 30} = c$$
4. $$\frac{r'}{n' + 40} = d$$
5. $$\frac{r' + 4}{n' + 40} = e$$

Then, for each equation $r'$ was replaced by $p*n'$ to obtain a value of $n'$ (hypothetical sample size). Accordingly, the following equations resulted:
(1) \( p*n' = a(n'+20) \)
(2) \( p*n'+2 = b(n'+20) \)
(3) \( p*n'+1 = c(n'+30) \)
(4) \( p*n' = d(n'+40) \)
(5) \( p*n'+4 = e(n'+40) \)

For each equation, an \( n' \) value could be obtained and there were five values of \( n' \) for all the questions. Then, the mean of these values of \( n' \), \( n* \), was computed, and the quartiles obtained were those of the beta distribution with parameters \( r* \) and \( n* \), where \( r* = p*n* = (r'/n')n*. \)

Of course, the values of quartiles were obtained or approximated from Schlaifer's cumulative beta fractiles table.

Quartiles of the prior distributions assessed by using the discrete PDF method were obtained by imputing the approximate cumulative probability densities of the histogram drawn for the intervals of various error rates; 0-2%, 2-4%, 4-5%, 5-6%, 6-8%, and greater than 8%.

For both cases A and B, average discrepancies were obtained at quartile values. These average discrepancies were divided by respective quartile values to

---

The procedure of determining the values of \( r' \) and \( n' \) described in Winkler's article (1967) is not in detail. Detailed explanations were provided to the experimenter by Professor Winkler in a private discussion session on this subject.
TABLE 4
DISCREPANCIES BETWEEN QUARTILES OF TWO PRIOR DISTRIBUTIONS
(THE EPS AND DISCRETE PDF METHODS)

<table>
<thead>
<tr>
<th></th>
<th>Case A</th>
<th></th>
<th>Case B</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>First Quartile</td>
<td>Median</td>
<td>Third Quartile</td>
<td>First Quartile</td>
</tr>
<tr>
<td>(1) Average discrepancy</td>
<td>.009</td>
<td>.008</td>
<td>.013</td>
<td>.006</td>
</tr>
<tr>
<td>(2) Average beta quartile</td>
<td>.034</td>
<td>.045</td>
<td>.057</td>
<td>.019</td>
</tr>
<tr>
<td>(3) Average discrepancy as a % of the beta quartile</td>
<td>26%</td>
<td>17%</td>
<td>23%</td>
<td>31%</td>
</tr>
<tr>
<td>(4) Greatest difference</td>
<td>.039</td>
<td>.044</td>
<td>.049</td>
<td>.031</td>
</tr>
</tbody>
</table>
TABLE 5
DISCREPANCIES BETWEEN QUARTILES OF TWO PRIOR DISTRIBUTIONS
(THE HFS AND DISCRETE PDF METHODS)

<table>
<thead>
<tr>
<th></th>
<th>Case A</th>
<th></th>
<th>Case B</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>First Quartile</td>
<td>Median</td>
<td>Third Quartile</td>
<td>First Quartile</td>
</tr>
<tr>
<td>(1) Average discrepancy</td>
<td>.008</td>
<td>.010</td>
<td>.019</td>
<td>.007</td>
</tr>
<tr>
<td>(2) Average beta quartile</td>
<td>.037</td>
<td>.049</td>
<td>.066</td>
<td>.016</td>
</tr>
<tr>
<td>(3) Average discrepancy as a % of the beta quartile</td>
<td>22%</td>
<td>20%</td>
<td>28%</td>
<td>43%</td>
</tr>
<tr>
<td>(4) Greatest difference</td>
<td>.031</td>
<td>.040</td>
<td>.058</td>
<td>.032</td>
</tr>
</tbody>
</table>
measure relative discrepancies for each case. In both cases, average discrepancies as percentages of beta quartiles were greater than Felix's results, but smaller than Corless's results. Certainly, limited training had some positive effect on the resulting discrepancies in Felix's study - decreased inconsistencies. As compared to Corless's study, it is believed that the use of the EPS and HFS methods made it possible for the subject-auditors to express their beliefs on the estimated error rates more consistently, because the experimental design in this study was similar to Corless's with the only clear difference being the methods used. It is possible, however, that the greater discrepancies in Corless's study were inflated due to some poor quality assessments made by a few subjects as indicated in the greatest discrepancies reported in Corless's study. Based on the results obtained in this study, it seems highly unlikely that there exists an error rate estimated at about 37 percent as compared to the average median value of about 4 percent unless there is a very 'careless' assessor.

There was no significant difference between the two comparisons; EPS-PDF and HFS-PDF. The inconsistencies

---

86See Corless, 1972, p. 561(Table 4).
observed for EPS-PDF and HFS-PDF were almost at the same level as reported in Tables 4 and 5. This result combined with the result observed in the previous part on auditors' perceptions of hypothetical samples methods strongly suggest that the EPS and HFS methods are compatible with auditors' backgrounds and with the audit environment.

In most of the assessments, the least discrepancies were found between median values. One possible explanation for this fact is that subjects think in terms of median values or mean values when they make judgments on the error rates resulting from the weakness of an internal control system. Graphs of cumulative probability density functions of a typical subject (identification number 8 from the Buffalo group) given in Figure 2 illustrate this phenomenon clearly. In virtually all the cases, the least discrepancies were observed for median values.

Locations of median responses. To investigate the amount of agreement among the subjects the possible error rates, differences in locations of median ($F_{.50}$) responses were obtained and are presented in Table 6.

A high degree of agreement was observed among median responses of the subject-auditors. One possible explanation of this agreement is that in the two cases reasonably detailed information on previous years' sampling results was provided and the auditors might
FIGURE 2

GRAPHS OF CUMULATIVE DISTRIBUTION FUNCTIONS
FOR SUBJECT #8
TABLE 6
MEDIAN RESPONSES FALLING WITHIN EACH ERROR RATE INTERVAL

<table>
<thead>
<tr>
<th>Error Rate Interval*</th>
<th>Case A</th>
<th>Case B</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EPS</td>
<td>HFS</td>
</tr>
<tr>
<td>0 - .01</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>.01 - .02</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>.02 - .03*</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>.03 - .04</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>.04 - .05</td>
<td>16</td>
<td>12</td>
</tr>
<tr>
<td>.05 - .06</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>.06 - .07</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>.07 - .08</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>.08 -</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

*Does not include upper limit for each interval.
have relied somewhat heavily on the previous sampling results in their estimate of the current period's error rates. The auditors, also, might have a reasonably uniform range of error expectations for a certain circumstance. On the other hand, significant differences were found in some subjects' estimates of median responses for the same case because of the use of different SPD assessment techniques reflecting inconsistency in their judgments as discussed previously.

Fifty percent credible intervals could not be tabulated, because there were so many overlapping ranges over two adjacent intervals. For example, the fifty percent credible interval of subject 1 (for Case A) ranged from .03 to .06 while that of subject 2 ranged from .02 to .05. Therefore, no meaningful classification of intervals could be made.

**The Effect of Background Factors on Discrepancies**

In an attempt to investigate the effect of personal background factors on discrepancies in assessed SPD and to find basic qualifications necessary for reasonably consistent assessments of SPD, eight simple bivariate regressions and two multiple regressions were run. The results of the bivariate analysis are presented in Table 7 and those of the multiple regressions are provided in Table 8.
### TABLE 7

**RESULTS OF SIMPLE BIVARIATE REGRESSIONS OF BACKGROUNDS ON DISCREPANCIES BETWEEN MEDIAN RESPONSES (Case A)**

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Explanatory Variable</th>
<th>Regression Intercept</th>
<th>Beta</th>
<th>Beta t-value</th>
<th>F-value</th>
<th>Significance</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_1$</td>
<td>$X_1$</td>
<td>.0128</td>
<td>-.0008</td>
<td>-1.308</td>
<td>1.638</td>
<td>.21</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td>$X_2$</td>
<td>.0227</td>
<td>-.0051</td>
<td>-2.036</td>
<td>4.109</td>
<td>.05</td>
<td>.32</td>
</tr>
<tr>
<td></td>
<td>$X_3$</td>
<td>.0117</td>
<td>-.0001</td>
<td>-1.780</td>
<td>3.170</td>
<td>.09</td>
<td>.20</td>
</tr>
<tr>
<td></td>
<td>$X_4$</td>
<td>.0322</td>
<td>-.0058</td>
<td>-1.301</td>
<td>1.527</td>
<td>.22</td>
<td>.21</td>
</tr>
<tr>
<td>$Y_2$</td>
<td>$X_1$</td>
<td>.0217</td>
<td>-.0019</td>
<td>-1.690</td>
<td>4.100</td>
<td>.05</td>
<td>.26</td>
</tr>
<tr>
<td></td>
<td>$X_2$</td>
<td>.0207</td>
<td>-.0037</td>
<td>-1.509</td>
<td>3.012</td>
<td>.07</td>
<td>.16</td>
</tr>
<tr>
<td></td>
<td>$X_3$</td>
<td>.0156</td>
<td>-.0002</td>
<td>-1.308</td>
<td>3.184</td>
<td>.08</td>
<td>.19</td>
</tr>
<tr>
<td></td>
<td>$X_4$</td>
<td>.0362</td>
<td>-.0063</td>
<td>-.689</td>
<td>1.341</td>
<td>.25</td>
<td>.26</td>
</tr>
</tbody>
</table>

**Note**

$Y_1 = EPS_F^{.50} - PDF_F^{.50}$

$Y_2 = HFS_F^{.50} - PDF_F^{.50}$

$X_1 = $ Audit experience

$X_2 = $ Statistical background

$X_3 = $ Statistical sampling instruction

$X_4 = $ Client size

For more details, refer to Methodology for purpose (3).
<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Explanatory Variables</th>
<th>Beta</th>
<th>F-value</th>
<th>Significance</th>
<th>Beta</th>
<th>F-value</th>
<th>Significance</th>
<th>Beta</th>
<th>F-value</th>
<th>Significance</th>
<th>Beta</th>
<th>F-value</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Y_1 )</td>
<td>( X_1 )</td>
<td>-.0004</td>
<td>.276</td>
<td>.60</td>
<td>-.0050</td>
<td>13.597</td>
<td>.01</td>
<td>-.0000</td>
<td>.792</td>
<td>.38</td>
<td>-.0019</td>
<td>.804</td>
<td>.37</td>
</tr>
<tr>
<td></td>
<td>Regression intercept</td>
<td>.0351</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ANOVA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regression</td>
<td>4</td>
<td>.0015</td>
<td>.00037</td>
<td>6.55</td>
<td>.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>31</td>
<td>.0018</td>
<td>.00006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coeff. of variability</td>
<td>94%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( R^2 )</td>
<td>.45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( Y_2 )</td>
<td>( X_1 )</td>
<td>-.0018</td>
<td>7.065</td>
<td>.01</td>
<td>-.0041</td>
<td>10.449</td>
<td>.01</td>
<td>-.0000</td>
<td>.193</td>
<td>.66</td>
<td>-.0015</td>
<td>.546</td>
<td>.46</td>
</tr>
<tr>
<td></td>
<td>Regression intercept</td>
<td>.0399</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ANOVA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regression</td>
<td>4</td>
<td>.0017</td>
<td>.00043</td>
<td>8.56</td>
<td>.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>31</td>
<td>.0016</td>
<td>.00005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coeff. of variability</td>
<td>70%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( R^2 )</td>
<td>.52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
There was no single background factor which accounted for most of the variances in the discrepancies in SPD assessments as explained by low coefficients of determination. In all categories, however, slightly negative relationships were found between each personal background factor and the discrepancy. This indicates that experience in auditing and sampling, and statistical knowledge enhances consistency in SPD assessments. The most important element of the auditors' backgrounds which affected the consistency in assessments was the subjects' statistical background expressed in number of courses they had taken in graduate and undergraduate programs. A very unique relationship was observed between the discrepancies $Y_2$ (median difference between HFS and PDF) and auditors' amount of actual experience in auditing, because no significant effect was found for the same background factor on the discrepancies $Y_1$ (median difference between EPS and PDF). A possible explanation for this result is that more experienced auditors could use the HFS method more confidently. However, because of the small sample size and some subjects' unusually strong backgrounds (15 years of audit experience, 6 courses in statistics, about 100 hours of sampling instruction, etc.), drawing a general conclusion in this respect is hazardous.
The size of the majority of the subjects' clients had virtually no impact on the resulting consistency, at least as compared to the companies A and B in the case. Subjects' sampling knowledge expressed in number of hours of instruction in statistical sampling they had received from their current and previous firms was found to have an effect on consistency to some extent. Again, in this category, some subjects had substantial amount of training and some did not have any previous training in audit sampling, and no strong conclusion is drawn.

Multiple regressions were run to simultaneously investigate the impact of various background factors on consistency. Results which were consistent with those of bivariate regressions were observed. Explanatory variable $X_2$, statistical background, displayed the most significant effect on consistency in the assessments, as was observed in the simple regressions. This effect completely swamped some minor effect observed of $X_3$, sampling knowledge. In the multiple regressions also, $X_1$, audit experience, had a significant effect on $Y_2$, median discrepancy between HFS and PDF, implying that more experienced auditors could use the HFS method more confidently.

Except indicating that a good statistical background and actual audit experience help auditors
assess SPD in a more consistent manner, the results did not suggest any minimum qualifications necessary for an auditor to be able to use the tested methods in audit tests. However, the auditor could use all three techniques (EPS, HFS, nad PDF) or any combination of two methods, check resulting discrepancies at each beta quartile, and compare them to the average discrepancies expressed as percentages of beta quartiles to test whether his discrepancies are above or below the reported averages. In this manner, the auditor could at least compare his assessments to "average" assessments if no "standard" consistency measures are obtainable, as is and will be in this discipline.

Compatibility of HFS with EPS

In the review sessions with four auditors for further refinement of the initial questionnaire, and from subject-auditors' post-response comments, it was apparent that the subjects interpreted the HFS questions as an extension of the EPS method. Many subjects commented that those hypothetical future samples questions made them reconsider how well the previously estimated proportion of r/n represented their true beliefs. Some subjects stated that after they completed the second questionnaire section (HFS part) they went back to the first part (EPS) and modified the initial assessment of r/n. Two subjects
viewed the HFS method as a formalized "stop-and-go" sampling method frequently used in current practices. Other various comments and the ratings assigned to the EPS and HFS methods indicated strongly that the auditor-subjects viewed these two hypothetical samples methods as two appealing methods and compatible with each other. The EPS question, in turn, provided a proforma assessment of the error rate to the subjects which could be used as a sound starting point for the HFS method.

This compatibility is a crucial characteristic in recommending the two hypothetical samples techniques as the most appropriate methods to be used by auditors, in addition to the appealing nature of the two methods to the auditors. Inconsistencies, if not serious enough to nullify the entire assessments, should not present a major problem, since inconsistencies can be removed if the assessor is made aware of their existence.\footnote{Winkler and Hays, 1975, p. 484.} In this regard, Winkler and Hays suggest that a SPD assessor attempt to assess the same set of probabilities in more than one way in order to check his consistency.\footnote{Winkler and Hays, 1975, p. 484.} Therefore, by using these two compatible methods, the consistency check could also be done to achieve a good quality assessment.
When an assessor uses the HFS method, however, some cautions should be kept in mind. Those hypothetical sample sizes should be of varied magnitude in order to provide a wide range of estimates. The hypothetical sample results expressed in the number of occurrences in a certain number of trials should also be designed to be flexible so that sufficient "challenges" would be made to extract meaningful responses.
SUMMARY, CONCLUSIONS, AND SUGGESTED FUTURE RESEARCH

Objectives Restated

To apply the SAS No. 1 formula for determining the extent of substantive testing and to use the Bayesian method of statistical sampling, the auditor's subjective evaluation of the internal control system of a client should be expressed in the form of a probability distribution. Auditors, however, are generally not statisticians, and there should be an appropriate method(s) of expressing this subjective evaluation for auditors with modest statistical backgrounds. A study of the appropriate SPD assessment technique(s) which would not overwhelm the auditor's modest background in probability and statistics is essential.

The primary purpose of this study was to test two SPD elicitation techniques which seem to be most congruent with the audit environment and auditors' backgrounds, and to investigate the basic qualifications which are necessary for auditors to be able to assess their prior distributions in a reasonably consistent manner. This primary purpose provided the general framework for subsequent research and the following specific purposes were subse-
quentiy defined:

1. To investigate the auditor's perception and acceptance of hypothetical samples techniques (EPS and HFS) of SPD elicitation.

2. To investigate the consistency in SPD assessments resulting from the use of two hypothetical samples techniques.

3. To investigate the effect of personal background factors on discrepancies in assessed SPD, and to find basic qualifications necessary for reasonably consistent assessments of SPD.

These issues were investigated using a research methodology which included a literature search, an in-depth review and analysis of three previous studies in this area, a questionnaire design, a survey of experienced auditors, and an analysis of the survey results. The results of the study and the conclusions drawn are summarized in this chapter.

Summary and Conclusions

1. The appealing nature of the two hypothetical samples techniques were confirmed. The EPS and HFS methods were perceived as easier than the direct assessment technique in both contexts; easy to understand and easy to use. Of the two techniques, the HFS method was rated as an
easier technique to understand than the EPS method. Also, in the how-easy-to-use context, the HFS method was rated lower (easier) than the EPS method, and the difference in the mean ratings was even greater as compared to the how-easy-to-understand context.

Simultaneous multiple comparisons indicated that there exists significant difference in the subjects' perceptions of the three different methods. The differences (in the perceptions) between the EPS method and the PDF method and between the HFS method and the PDF method were significant, while there was virtually no difference between the EPS and HFS methods.

No significant differences were found in the analysis of the ratings of a non-auditor group as to the three methods, indicating that non-auditors did not perceive any appealing nature of the two hypothetical samples techniques. Thus, it was concluded that the only distinguishable characteristic of the auditor group, actual audit experience, as compared to the non-auditor group, played a major role in making the two hypothetical samples techniques appealing to the auditors. Some difficulties in using the EPS method were experienced by the non-auditors.

2. In both cases (A and B), average discrepancies between quartile values of two prior distributions obtained
by using the EPS method and the discrete PDF method (and two distributions obtained by using the HFS method and the PDF method) as percentages of beta quartiles (from EPS and HFS) were greater than Felix's results, but smaller than Corless's results. Certainly limited training had some positive effect on resulting discrepancies in Felix's study and decreased inconsistencies in subjects' assessments.

As compared to the Corless study, it is believed that the use of the EPS and HFS methods made it possible for the subject-auditors to express their beliefs on the estimated error rates more consistently, because the experimental design in this study was similar to Corless's with the only clear difference being the methods used. It is possible, however, that the greater discrepancies in Corless's study were inflated due to some poor quality assessments.

There was no significant difference between two comparisons; EPS-PDF and HFS-PDF. The inconsistencies observed for EPS-PDF and HFS-PDF were almost at the same level. This result combined with the result observed in the previous part on auditors' perceptions of hypothetical samples methods strongly suggest that the EPS and HFS methods are compatible with auditors' backgrounds and with the audit environment.
In most of the assessments, the least discrepancies were found between median values. Subjects may have been thinking in terms of median values or mean values when they made judgments on the error rates resulting from weaknesses of an internal control system. This may indicate that naive assessors feel more comfortable with point estimation than with distribution estimation.

A high degree of agreement was observed among median responses of the subject-auditors. One possible explanation of this agreement is that the case used reasonably detailed information on previous years' sampling results and the auditors might have relied heavily on the previous sampling results in their estimate of the current period's error rates. The auditors also might have a reasonably uniform range of error expectations for a certain circumstance. On the other hand, significant differences were found in some subjects' estimates of median responses for the same case because of the use of different SPD assessment techniques reflecting inconsistency in their judgments.

3. There was no single background factor which accounted for most of the variances in the discrepancies in SPD assessments. In all categories, however, slightly negative relationships were found between each personal background factor and the discrepancy. This indicates
that experience in auditing and sampling, and statistical knowledge enhances consistency in SPD assessments.

The most important element of the auditors' backgrounds which affected the consistency in assessments was the subjects' statistical background expressed in number of courses they had taken in graduate and undergraduate programs. A very unique relationship was observed between the discrepancies between median responses of HFS and PDF and auditors' amount of actual audit experience, because no significant effect was found of the same background factor on the difference between median responses of EPS and PDF. A possible explanation for this result is that more experienced auditors could use the HFS method more confidently. However, because of the small sample size and extremely varied backgrounds of some subjects, drawing a general conclusion in this respect is hazardous.

The size of the majority of the subjects' clients had virtually no impact on the resulting consistency, while the subjects' sampling knowledge was found to have some effect on consistency. When the impact of various background factors on the consistency was analyzed simultaneously, the effects of the statistical background and audit experience swamped other minor effects. Except for indicating that a good statistical background and actual audit experience help auditors assess SPD in a
more consistent manner, the results did not suggest any minimum qualifications necessary for an auditor to be able to use the tested methods in audit tests.

The auditors could use any combination of two methods, check resulting discrepancies at each beta quartile, and compare them to the average discrepancies expressed as percentages of beta quartiles to test whether his discrepancies are above or below the reported averages. In this manner, the auditor could compare his assessments to "average" assessments in the absence of any "standard" consistency measures.

4. The results of the survey also indicated that the subjects interpreted the HFS questions as an extension of the EPS method. Comments indicated that those hypothetical future samples questions made the subjects reconsider how well the previously estimated proportion of r/n represent their true beliefs. Two subjects viewed the HFS method as a formalized "stop-and-go" sampling method frequently used in current practices. It is concluded that the auditor-subjects viewed the two hypothetical samples methods as two appealing methods and compatible with each other. The EPS question, in turn, provided a proforma assessment of the error rate to the subjects which could be used as a sound starting point for the HFS method.
This compatibility is a crucial characteristic in recommending the two hypothetical samples techniques as the most appropriate methods to be used by auditors, in addition to the appealing nature of the two methods. Inconsistencies should not present a major problem, since inconsistencies can be removed if the assessor is made aware of their existence. By using these two compatible methods the consistency check could also be done to achieve a good quality assessment.

Suggested Future Research

Future research in this area should involve (1) training, (2) assessment, (3) feedback, and (4) reassessment phases in order to provide more working guidelines to the auditing profession regarding the application of SPD in evaluating the reliability of internal control systems for audit decision making.

1. **Training.** Even limited training in probability, especially implications of various magnitudes of beta parameters on the shapes of resulting probability distribution functions, would increase the consistency in assessments. It would be worthwhile to observe the subject-auditors' perceptions of two hypothetical samples techniques prior to training and subsequent changes in perceptions after training. Realistically, training in SPD
assessment for an extended period of time is not feasible, because obtaining a reasonable number of subjects and keeping the personal interests of the subjects in a positive direction would be very costly, if not impossible. A feasible alternative is to provide the interested subject-auditors an easy-to-read manual in SPD assessment including an explanation of the impact of the estimated beta parameters (r and n in the previous conjugate prior distribution discussion) on the shapes of the PDF's about two weeks before the scheduled assessment session, so that subjects could get themselves familiarized with SPD. A brief session to review SPD techniques and concepts would be helpful prior to the assessment session.

2. **Assessment.** The two compatible hypothetical samples methods have both a priori and sufficient empirical evidence to be the most appropriate methods in an audit environment. The EPS questionnaire format used in this study can be used in any error rate estimation for audit attributes without any modification. The HFS questions, however, should be made up for each particular audit situation based primarily on the previous audit sample results and changes in the internal control system being tested. There are two requirements in this task: One, the hypothetical sample sizes should be large enough to affect the assessor's initial assessment of the error
rate. If the initial estimate is not modified, that question cannot be used in computing the parameter $n$ for the conjugate beta distribution. Two, the hypothetical proportions, $r/n$, should be of wide variety to elicit a flexible estimate of the assessor's judgment on the error state.

3. **Feedback.** There is no single correct answer in SPD assessment. However, any existing inconsistency in the assessor's judgment which may be the result of vagueness, carelessness, or incomplete understanding of the method of eliciting SPD, could be reduced or eliminated by a mere awareness of the existence. Due to the limitations in administration, an immediate feedback to the initial assessment may be difficult to provide to the subjects in future experiments also.

4. **Reassessment.** If the feedback and reassessment could be performed in interview sessions, a better implementation of the planned experiment could be done. Realistically, a second interview with the same subject-auditor may be difficult to schedule, and the inconsistencies in the initial assessments could be informed to the subjects by mail. A second type of questionnaire for the reassessment should be designed and provided to the subject for determining which distribution represents his true belief. If this phase of the experiment is well
executed, a much more accurate assessment result could be obtained. Ultimately, for an actual application of SPD assessment to audit sampling and to a decision theoretic approach to audit decision making, this reassessment procedure should be included in the training.
BIBLIOGRAPHY
BIBLIOGRAPHY


APPENDIX A

THE QUESTIONNAIRE
About the Experiment

The second standard of field work requires that the auditor study and evaluate client’s system of internal control to determine the extent of substantive tests for a certain desired level of assurance. To use SAS A1 formula for the reliability level for substantive tests, the reliance assigned to internal control should be expressed by the auditor in the form of probability.

In statistical sampling for both tests of transactions and substantive tests, the sample size could be significantly decreased while maintaining the same level of reliability by using another statistical technique called the Bayesian method.

For both of the above applications, the auditor’s subjective evaluation of internal control should be expressed in the form of a probability distribution. However, auditors are generally not statisticians, and there should be an appropriate method of expressing this subjective evaluation for auditors with modest statistical backgrounds. (In the national offices of two New York based “big Eight” firms, auditors are now asked to specify audit risk level in the form of probability.)

There are various alternative methods of quantifying the auditor’s judgment available from previous psychological researches. To find the most appropriate method of the alternatives, a test of these methods with experienced auditors as subjects is essential.

Following is an audit case and related error estimation which can be done by responding to the indirect questions made up for this task. In the entire questionnaire, we are only concerned about the number of sales invoices in error (regardless of the number of errors in one invoice).

Assume that today is February 15, 1978 and you are auditing FY 1977 figures.
Texas Hardware Co. is a small wholesale distributor of hardware to independent, high-quality hardware stores in Louisiana and Texas. The audited 1976 figures are: net sales $1,650,401, total assets $580,162, and net income after tax $44,409. This is the third year of the audit for this client since 1975, and there were no significant errors discovered in the tests in the 75 and 76 audits performed by other colleagues. The president, Richard Chulick, has been the chief operating officer for approximately 10 years. He is regarded as a highly competent, honest individual who does a conscientious job.

The following information is provided from the auditor's files:

**Organization Chart - Personnel**

- **President** (not an owner)
- **Owners/Directors**
- **Sales**
- **Purchasing/Inventory Control**
- **Chief Accountant**
- **Order Clerk**
- **Billing Clerk**
- **Bookkeeper**

**Flowchart of Order-Receiving/Shipping/Billing**

Order Clerk → Shipping Dept → Billing Clerk → Bookkeeper

Sales recorded
Notes to the flowchart:

1. All sales order numbers and all bills of lading numbers are accounted for weekly by the accountant.
2. Sales account recorded on sales invoice is based on standard price list.
3. Statements are sent to customers monthly.

(OTHER CONTROL FEATURES FOR OTHER FUNCTIONS OMITTED)

Last year audit procedure for billing function

1. Trace selected shipping documents to a duplicate sales invoice for assurance that each one has been billed. Audit result: No error.
2. Trace selected duplicate invoice numbers from the sales journal to duplicate sales invoice, and check the accuracy of sales invoice preparation. (This procedure was performed, for convenience, together with the audit test of bill of lading, duplicate sales order, and customer order, at the same time because supporting documents are all attached to the duplicate sales invoices.

**Statistical Sampling Data Sheet - Attribute** (Year end 12/31/76)

Client: Texas Hardware
Audit area: Tests of transactions - Billing function
Objective: Examine duplicate sales invoices and related documents to determine if the system is functioning as intended.
(Sample size for the same period in 77: 5,120)
Sampling unit: Sales invoice
Organization of population items: Recorded in the sales journal sequentially.
Random selection procedure: Random number table

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Sample size</th>
<th>No. of erroneous invoices</th>
<th>Sample error rate</th>
<th>Nature of Exceptions</th>
<th>Effect on the Audit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy of sales invoice</td>
<td>150</td>
<td>9</td>
<td>6%</td>
<td>In 4 cases the wrong price was used but the errors were under $10 each case.</td>
<td>Test the 50 largest sales transactions for the year for proper price and expand the confirmation of accounts receivable.</td>
</tr>
<tr>
<td>preparation</td>
<td></td>
<td></td>
<td></td>
<td>In 3 cases freight was not charged.</td>
<td>(Results: No material exceptions)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>In 2 cases there was an extension error of $1,000 and $100.</td>
<td></td>
</tr>
</tbody>
</table>

For FY 1977, some improvements in internal accounting control were made. The improvement made with regard to the attribute being studied in this case is that:

Duplicate sales invoices are compared with bill of lading, and prices are checked daily by the order entry clerk, and initialed before the original invoice is mailed to the customer. (where was no internal verification before)

This verification is Mike Ilia's additional duty. There was no personnel increase due to this new control procedure. Unaudited 1977 figures: Sales $1,540,090

Total assets $509,765

Net income after tax $41,928

**************************************************

Based on your own audit experience what would be the percentage of the number of invoices in error after the improvements in internal control in 1977? (Note: 6% error rate in 1976)

Of course you are not sure as to the exact percentage error. By answering the following questions, you can assess your own probability.

A) Determine two numbers (r) and (n) such that your own error rate estimate would be roughly equivalent to having observed (r) invoices in error in a random sample of (n) sales invoices in any of your past audits similar to this. r = \( \frac{r}{n} \) This is not the final audit sample size. Of course (n) does not have to be 150 which was the sample size another auditor used last year. This should be your own decision.

B) How do you estimate the error rate? ( )%. Please think in terms of %. (The proportion r/n computed in (A) would be helpful in doing this). Please imagine carefully how you would feel if presented with each of the following random samples. Each sample is independent, so consider what effect each sample by itself would have on your original probability.

(a) A random sample of 20 sales invoices - No error. Now how would you adjust your original probability? (i.e., error rate 6%, 5%, etc.) ( )

(b) Sample 20 invoices - 2 errors. Now how would you adjust? ( )%

(c) Sample 30 invoices - 1 error. Now how would you adjust? ( )%

(d) Sample 40 invoices - No error. What is your adjusted rate? ( )%

(e) Sample 40 invoices - 4 errors. Now how would you adjust? ( )%

C) If you were auditing 100 companies similar to Texas Hardware (with the same set of circumstances), how many of them would you expect to have actual error rates which are contained in each of the following intervals:

- less than 2% ( ) companies
- 2% - 4% ( ) companies
- 4% - 5% ( ) companies
- greater than 6% ( ) companies

Total 100 companies

Audit Case B

Greater Louisiana Hardware Co. is a medium-size wholesale distributor of hardware. In overall size, this company is ten times as large as the company in case A, in all categories. As to organization, instead of single bookkeeper, order clerk, and billing clerk, separate departments perform respective functions (credit manager, billing department, central files, sales order department, cashier, accounts receivable department, etc.)
Internal control features for the verification of sales invoice preparation

1) Sales prices and credit terms are based on approved standard price lists. Any deviations are approved by sales manager.

2) A designated billing clerk receives the shipping advices directly from the shipping department.

3) Sales invoices are checked as to prices, extensions, and footings against customers' orders and shipping advices by the office manager.

(For other features of internal control, please assume typical situations for a firm of this size and organization, based on your experience.)

Now, how would you assess the error in sales invoice preparation for the company in case B? (There is not enough information for you to do this comfortably. Please give answers by adjusting your answers in case A)

A) 

<table>
<thead>
<tr>
<th>( )</th>
<th>( )</th>
<th>( )</th>
<th>( )</th>
<th>( )</th>
<th>( )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
<tr>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
<td>( )</td>
</tr>
</tbody>
</table>

Total: 100 companies

1) How many years of audit department experience do you have? ( )
2) How many statistics courses did you have at the college (undergraduate and grad.) ( )
3) How much instruction in statistical sampling have you received from your firm (current and previous)? ( )
4) How much outside continuing education have you had in audit sampling? ( )
5) How would you compare the size (sales) of the majority of your clients with that of company A and company B? Bigger than B ( ) Similar to B size ( )
6) For case B, when you were estimating the error, how did you assume the internal control to be? Strong internal control ( ) Moderate ( ) Weak ( )
7) How easy was it to understand each method before you used the method?

First method: 
Very easy ( ) Moderate ( ) Very difficult ( )
Second method: 
Very easy ( ) Moderate ( ) Very difficult ( )
Third method: 
Very easy ( ) Moderate ( ) Very difficult ( )
8) How easy was it to use actually?

First method: 
Very easy ( ) Moderate ( ) Very difficult ( )
Second method: 
Very easy ( ) Moderate ( ) Very difficult ( )
Third method: 
Very easy ( ) Moderate ( ) Very difficult ( )
In the October 1976 issue of The Accounting Review, Professor Felix published the results of an experiment on the feasibility of the Equivalent Prior Sample (EPS) method of Winkler (1967) to assess subjective probability distributions for audit decisions. The purpose of his research was "to give additional exposure to" the EPS method which "appears to be particularly congruent with the auditor's environment." His research results indicated that the percentage differences based on the EPS method were "somewhat smaller" than the differences of the results obtained by Corless (1972) using the direct assessment of quartiles method. For the attributes 1 and 2, the average difference as a percentage of the beta quartile ranged from 9 percent to 32 percent as compared to the 25-45 percent range reported by Corless. While the basic idea of viewing qualitative audit evidence as equivalent to a prior sample should be appealing to auditors because of its relative easiness and understandability, there is a weakness in his experiment which might bring a question on the more consistency the EPS method could yield.

Felix used as subjects ten practicing auditors which "were obtained on the basis of availability and interest." He did not attempt to control demographic
factors, such as the level of education in statistics and probability or the amount of experience in audit sampling, "since they had similar backgrounds." Felix apparently thought there would not be any significant effect on consistency as long as subjects had similar backgrounds, based on Winkler's statement that "education and training in statistics can be a significant factor when some subjects have extensive backgrounds and others do not." However, when these subjects have the same low level of training in probability assessment, the task of assigning the quartiles becomes a very difficult one resulting in relatively inaccurate, inconsistent subjective probability assessments. The probability distributions will naturally have more dispersion as compared to the case of the direct fractiles assessment method by sophisticated assessors or the case of the EPS method used by relatively inexperienced assessors. "Some" difficulty in using the direct fractiles method was experienced by "all" of the auditors in the Felix experiment and one auditor (10% of the Felix' sample size) thought the first and third quartiles as the location of a 50 percent credible interval centered on the median. Accordingly, it is not very hard to understand that an easier method (EPS) will produce less inconsistent assessments.

It should be interesting to note here an evaluation by some British scholars. Hampton, Moore and Thomas (H, M & T) (1973, p. 30) concluded, after a critical analysis
of the literature on subjective probability and its measurement, that the direct fractiles assessment method seems to be the method most likely to achieve the end of obtaining consistent assessments. It is true that there should be some sacrifice of theoretical sophistication for a more basic, easy-to-apply approach. However, an actual application of the Bayesian approach to audit decisions requires a fairly high degree of knowledge and training in probability and statistics on the part of the auditors, because, unless the quality of the assessment of subjective probability distributions is good, the benefit of using the Bayesian approach over the classical statistical technique cannot be justified. Corless (1975, p. 158) commented, based on his experiment results, that, until we can be sure of the quality of prior distributions assessed by the auditors, it would not be wise to use the priors, particularly when these priors conflict with the sampling evidence. Of course, until the results of an experiment with experienced and trained auditors as subjects indicate the firm evidence that the direct quartiles method results in more consistent assessments, we cannot accept the above statement of H, M & T for use in auditing environment. I don't see, however, any merits in trying to find the evidence that an easier method results in less inconsistency using the auditors with very limited training as subjects.
Felix suggested in his conclusion the execution of training experiments with experienced auditors as subjects quoting the study of Schaefer and Borcheoding which indicated training had a considerable effect on the performance of assessors and the methods they use most effectively. The observation of any change in the consistency of assessments based on the direct fractiles method and the EPS method at different stages of training would be particularly interesting in relation to the above discussion on the method Felix used. So far, both Corless and Felix confined their studies to the individual assessments of subjective probability distributions. To alleviate errors from unnecessary human bias, the importance of interpersonal agreement (intersubjectivity) has been recognized for objective observations (The Committee on Basic Auditing Concepts, 1969-1971, p. 45). In addition to finding the method for individual assessment, an effort to try the group consensus approach would be worthwhile. Rather than the mathematical combination of individual assessments, an alternative approach to group consensus through the modification of individual assessments seems to be more congruent with audit environment. Above Winkler's paper considered two techniques: feedback and re-assessment (FR) and group re-assessment (GR). While the Delphi Technique which is a variant of Winkler's FR method is of considerable interest because of the reduction of the
influence of the dominant individual (Hampton et al., 1973, p. 32), GR method may also be included for further research based on the assumption that the most experienced auditor should influence less experienced ones in the judgment process, and the more experienced auditor's assessments would be more accurate, or at least, less inconsistent.
REFERENCES


APPENDIX C
FELIX'S RESPONSE
January 23, 1978

Mr. John Y. Lee
Louisiana State University
School of Business Administration
Baton Rouge, Louisiana 70803

Dear Mr. Lee:

I apologize for being so tardy in responding to your request for a comment on the note that you wrote on my article. It seems to me that the primary thrust of your comment is that the hypothetical future sampling or the HFS method suggested and tested by Winkler may be superior to the EPS method and should also be tested. For a number of reasons, including some other literature sighted in my article, I doubt that this method would be superior to the EPS method. But, I certainly do agree with the idea that it is worth testing. I chose not to do so in my study. In suggesting the use of the HFS method, you ought to be cautious in relying too heavily on the Winkler results. More recent studies have indicated that particularly after training this apparent advantage may not exist. Also, the advantage may not exist in different substantive studies.

Regarding your other comments, the primary thrust of your comments seems to be some unhappiness with my lack of consideration of the personal background of the subjects. I did collect biographical and educational data on my subjects, but found that they were essentially homogeneous in their training and background. For this reason I made no attempt to include them in the study. In particular, there was almost no difference in the amount of auditing experience or statistical background of the subjects.

In summary, I think it is fair to say that I have no particular quarrel with any of your comments. The major weakness of my study was the small sample size, ten subjects is really not enough, but given the purpose of my study, to study the feasibility of the EPS method and to suggest it for further research in auditing, I think
my objectives were reasonably met. I would predict that you are going to have some difficulty in getting a reviewer to view your comment as being significant enough for publication. One very obvious means of getting around the problem would be to carry-out an experiment with the HFS method on a reasonably good size sample of auditors. If you chose to go this route I would be very happy to provide you with any of the material I used to assist you in your experiment.

Cordially,

William L. Felix, Jr.
Associate Professor
of Accounting

WF:sk
VITA

Johng Yul Lee was born in Euisung, Korea on November 6, 1947. His elementary and high school education was completed in the Kyungbuk, Korea, public schools in 1965. He majored in Economics at Seoul National University, Seoul, Korea, receiving his Bachelor of Economics degree in 1970. While working on the undergraduate degree, he passed the first and second examinations for certified public accountant of Korea. He worked for the Chase Manhattan Bank from 1970 to 1974. From 1974 to 1975 he was a Rotary Foundation Fellow at Owen Graduate School of Management, Vanderbilt University, Nashville, Tennessee. He received the Master of Science degree in Accounting from Louisiana State University in 1977. While working on his Doctor of Philosophy degree in Accounting, he became a member of Beta Gamma Sigma and served as a Graduate Assistant in Accounting. He is currently Assistant Professor of Accounting at State University of New York at Buffalo, and a member of the American Accounting Association.
Candidate: Johng Yul Lee

Major Field: Accounting


Approved:

[Signatures]

Major Professor and Chairman

Dean of the Graduate School

EXAMINING COMMITTEE:

[Signatures]

Date of Examination:

April 24, 1979