80-11400

SMIELIAUSKAS, Waldemar John, 1947-
SIMULATION ANALYSIS OF THE STATISTICAL VALIDITY OF THE INTERNAL CONTROL HYPOTHESIS OF AUDITING WITH IMPLICATIONS FOR SUBSTANTIVE TESTING METHODS AND LINKAGE RULES.

The University of Wisconsin-Madison, Ph.D., 1980, Business-Mathematics

Xerox University Microfilms, Ann Arbor, Michigan 48106

(This title card prepared by the University of Wisconsin)

PLEASE NOTE:

The negative microfilm copy of this dissertation was prepared and inspected by the school granting the degree. We are using this film without further inspection or change. If there are any questions about the film content, please write directly to the school.

UNIVERSITY MICROFILMS

SIMULATION ANALYSIS OF THE STATISTICAL VALIDITY OF THE INTERNAL CONTROL HYPOTHESIS OF AUDITING WITH IMPLICATIONS FOR SUBSTANTIVE TESTING METHODS AND LINKAGE RULES

A thesis submitted to the Graduate School of the University of Wisconsin-Madison in partial fulfillment of the requirements for the degree of Doctor of Philosophy

BY

WALDEMAR JOHN SMIELIAUSKAS

Degree to be awarded: December 19____ May 19 80 August 19____

Approved by Thesis Reading Committee:

Paul H. Walgenbach
Major Professor

July 25, 1979
Date of Examination

Larry E. Rittenberg

Robert B. Miller

Robert M. Bock
Dean, Graduate School

SIMULATION ANALYSIS OF THE STATISTICAL VALIDITY OF THE INTERNAL CONTROL HYPOTHESIS OF AUDITING WITH IMPLICATIONS FOR SUBSTANTIVE TESTING METHODS AND LINKAGE RULES

BY

WALDEMAR JOHN SMIELIAUSKAS

A thesis submitted in partial fulfillment of the
requirements for the degree of

DOCTOR OF PHILOSOPHY
(Business)

at the

UNIVERSITY OF WISCONSIN-MADISON

1980

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

Figure Page

Figure Page

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to express my appreciation for the assistance, advice, and encouragement that I received during my Ph.D. program at the University of Wisconsin. I am particularly grateful to my dissertation committee composed of Professors Paul Walgenbach, Robert Miller, Larry Rittenberg, Averil Law, and Hugh Juergens.

Special mention should be made of the notable contributions by the three members of my reading committee. Professor Walgenbach's encouragement and continual support of my work served as a constant source of inspiration to me. Professor Rittenberg's thought provoking, critical analysis and feedback throughout the project contributed considerably to the development of the dissertation. And Professor Miller's patient statistical guidance greatly assisted the researcher's understanding of the statistical issues as well as aiding the research design. To all of these individuals I express my sincere gratitude and appreciation.

I am also grateful to the Peat, Marwick, Mitchell Foundation for their generous support which eliminated many resource constraints and enabled me to work full time on this dissertation. Finally, I wish to thank my typist, Holly Spiel. Her exceptional typing skills and concern for accuracy relieved much of the burden of the last stages of the dissertation.

To my wife Chantal whose sacrifice,
encouragement, and support made it
all possible.

CHAPTER ONE

Introduction

1.1 Preface

Recent trends in auditing research show an increasing interest in the problems of statistical sampling of auditing populations. This is probably the result of several factors at work in the business environment. These include the general growth in size and complexity of business operations and information processing, and the resultant need for more testing on the part of the auditor. Another factor is the prevailing litigious climate which is forcing auditors to adopt more objective measures of auditing tests. Yet another factor is the keen interest in cutting costs through the use of sampling theory to minimize the extent of testing. All of these factors provide persuasive arguments for considering statistical sampling issues an important area of auditing research.

This dissertation addresses a virtually unexplored aspect of statistical sampling in auditing--the relationship of the extent of substantive sampling to internal control information. More generally, it is concerned with the validity of the internal control hypothesis of auditing. That is, is it possible to rely on internal control information to reduce the extent of subsequent audit procedures and still maintain the reliability of the audit?

To some this initially may appear to be an unimportant question because the answer to the question is seemingly obvious. After all, this assumption is implied by the second standard of fieldwork and just about every textbook and theoretical work on auditing. In addition, the experience of auditors in the field appears to attest to its validity and audit firms are relying on this assumption to help reduce audit costs.[1] So one might wonder that with theory and practical experience appearing to support the hypothesis, why is there a need to analyze its validity through research? There are several reasons.

First, this hypothesis or assumption has never been tested formally, and, therefore, it remains exactly that--an assumption. Experience may not be confirming the validity of the hypothesis. Because of the lack of experimental control, it is difficult to measure how much risk auditing firms are actually experiencing or how efficiently audits are being conducted. Auditors may in fact be overtesting or undertesting. Informal experience alone has not indicated under what conditions the hypothesis holds nor to what extent it holds.

Second, even assuming the hypothesis is true, there is little in the way of theory to guide the auditor in deciding how much internal control information is cost-benefit justified. Practical experience has failed to specify under what environmental conditions and what combinations of audit methods this assumption holds and introduces the most efficiencies. As an example of this problem, consider a recent

[1]For a very recent example, read the statements made by Arthur Anderson's chairman, Harvey Kapnick, in "Holding the Line of Audit Fees," Business Week, (October 23, 1978), pp. 57-58.

study by Theodore Mock and Jerry Turner.[2] A key aspect of this study was to define normative statistical sample sizes for substantive testing based on the states of internal control used in the study. However, Mock and Turner apparently felt there was some ambiguity in deriving the normative sample sizes for the study, thus reflecting in part a deficiency of present audit research to provide guidance in the computation of optimal sample sizes under different internal control conditions. Furthermore, the fact that the subjects (all experienced auditors) in the study planned substantially different sample sizes from the predetermined normative sample sizes (Table 5 of their paper) provides additional evidence that the validity of such normative sample sizes needs to be investigated.

Finally, recent research has been casting doubt on the validity of the hypothesis. For example, one ramification of the Neter and Loebbecke Study is that since actual reliability can differ significantly from the nominal or stated reliability of statistical tests of accounting populations, auditors cannot be certain how reliance on internal control affects these actual sampling risks.[3] In addition some behavioral research has shown (in contrast to the results obtained by Mock and Turner, and others) that the amount of internal control

---

[2]Theodore J. Mock and Jerry L. Turner, "A Field Test of the Effect of Changes in Internal Controls in Audit Programs," a draft presented by Professor Mock at the University of Wisconsin-Madison on April 5, 1978.

[3]John Neter and James Loebbecke, Behavior of Major Statistical Estimators in Sampling Accounting Populations, with a Foreward by Paul Rosenfield, Auditing Research Monograph 2, New York, AICPA, 1975.

information has no effect on the subsequent audit plan. To quote from a conclusion of a recent study supported by the AICPA: "This research joins other research in questioning the nature of the relationship between evaluation of internal control and the substantive testing plan proposed."[4]

All of these reasons provide grounds for studying the validity of the internal control hypothesis. The question of this validity constitutes a major gap in auditing theory. Therefore if auditing theory is to be put on a more scientific basis, it is apparent that research should be directed to this topic; this provides the motivation for the present dissertation. (Note that the issue that is being raised here is not whether auditors do rely on internal controls [a behavioral research question], but whether they can rely on internal controls [an analytical research question]).

1.2 Statement of the internal control hypothesis

The basic goal of this dissertation is to test the statistical validity of what is referred to here as the internal control hypothesis of auditing. Although the hypothesis has never been formally stated as such, it is implied by auditing theory and by auditing standards.[5] For

---

[4]Ronald Weber, "Auditor Decision Making: A Study of Some Aspects of Accuracy and Consensus," (Ph.D. dissertation, University of Minnesota, 1977), pp. 188.

[5]The only explicit reference to an internal control hypothesis that the researcher has encountered is in Mock and Turner's "Effect of Internal Controls on Audit Programs," pp. 11.

example, the second standard of fieldwork is the following:

> "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"[6]

This standard is further discussed in Sec. 320A paragraphs 18 and 19:

> .18 The second standard of fieldwork requires that an evaluation of internal control as a basis for determining the extent of audit tests. [italics added] Compliance with this standard involves two problems: (a) evaluation of internal control, and (b) relating the extent of tests to this evaluation.
>
> .19 The second standard of fieldwork recognizes that the extent of tests required to constitute sufficient evidential matter under the third standard should vary inversely with the auditor's reliance on internal control. These standards taken together imply that the combination of the auditor's reliance on internal control and on his auditing procedures should provide a reasonable basis for this opinion in all cases. For statistical samples designed to test the validity or bona fides of accounting data and to be evaluated in monetary terms, the committee believes the foregoing concept should be applied by specifying reliability levels that vary inversely with the subjective reliance assigned to internal control and to any other auditing procedures or conditions relating to the particular matters to be tested by such samples. [The alternative ways of operationalizing this inverse relationship are referred to in this dissertation as linkage rules.][7]

There is thus a very strong implication that the subsequent audit procedures, which consist of substantive tests, can be reduced as a result of internal control information if that information indicates reliance is possible. There is in fact the stronger implication: that

---

[6]American Institute of Certified Public Accountants, Statement on Auditing Standards, Codification of Auditing Standards and Procedures No. 1. (New York: AICPA, 1973) Sec. 320A, (hereafter frequently cited as SAS No. 1).

[7]Ibid., Sec. 320A paragraphs .18 and .19.

the cost of obtaining internal control information is generally more than offset by the resultant savings in the subsequent audit work. However, as pointed out in section one of this chapter, recent research results cast doubt on the validity of these propositions. Therefore, because of the concern shown by the rise in auditing costs and the fact that internal control evaluation is required by the second standard of fieldwork, it is important to specify under what environmental conditions, if any, and under what combinations of audit methods these assumptions hold and introduce the most efficiencies. The dissertation is intended to be a first step in this direction. However, only the first implication, impact on extent of substantive testing, will be directly addressed in this dissertation since cost-benefit justifications require additional assumptions to be made about the costs of various tests. It is felt that such assumptions can only be made arbitrarily here and are best left to the individuals and firms actually performing these tests in practice.

In its most general form the internal control hypothesis pertains to all forms of audit testing, judgmental as well as statistical. However, since only statistical tests allow the objective control of risks associated with sampling, the validity of the internal control hypothesis must be tested within a statistical sampling framework. Nevertheless, this approach should also have some relevance to judgmental tests because the chief difference between statistical and judgmental tests is in the degree of objectivity in controlling the risks associated with sampling.

The basic goal of the dissertation, then, is to contribute to the development of auditing theory by examining the validity of the internal control hypothesis in a statistical sense, i.e., in a situation where all testing is done on a statistical basis and nonsampling risks have been eliminated.[8] The research will be made operational by using the following form of the internal control hypothesis:

> HO: Internal control information can be used to reduce the statistical sample size of substantive tests without increasing the actual audit risks that arise as a result of using an audit statistical sampling strategy.

It should be noted that this internal control hypothesis relates to the performance of what is called here the audit statistical sampling strategy and not just a substantive testing method. By audit sampling strategy is meant the combination of substantive testing method, amount of internal control information, and linkage rule (the rule used to relate internal control reliance to the substantive test sample size) used by an auditor to reach a statistical decision about an item on the financial statements. It is necessary to consider all three factors in defining a strategy because they all may impact on the validity of the internal control hypothesis. That is, the performance of a sampling strategy will be affected not only by the amount of internal control information, but also by the performance of the particular substantive sampling method and the form of the linkage rule used by the strategy. Hence an integrative approach is necessary to examine the validity of

---

[8]The impact of nonsampling risks may very well be the most important factor affecting the performance of an audit strategy. However, since virtually nothing is known about the form of audit nonsampling error distributions, this dissertation confines itself to exploring the potential impact by considering only one form of such distributions--the normal distribution. This is discussed in chapter four of the dissertation.

the internal control hypothesis. However, no one has yet taken such an approach even though this concept of sampling strategy more closely parallels the usual audit process as defined in Statement on Auditing Standards No. 1 (SAS No. 1) Sec. 320.[9]

It should also be noted that this form of the internal control hypothesis does <u>not</u> state that internal controls must always be relied upon--internal controls are only relied upon when the internal control information indicates reliability. What the hypothesis does state is that a strategy using such information over a whole range of internal conditions and accounting populations will have a smaller average substantive test sample size and actual audit risks comparable to or less than an audit strategy not using internal control information (an informationless strategy). This interpretation of the internal control hypothesis seems to be the one implied by the second standard of fieldwork.

Since there are two audit risks associated with each strategy: the risk of a Type I error (the error of rejecting a materially correct recorded amount) which is also frequently referred to as the α risk,

---

[9]Although William R. Kinney outlined and illustrated such an approach through his conception of routes to an opinion, he did not consider the diverse issues associated with a more realistic environment. In particular he implied that nominal audit risks always equal actual audit risks, that the auditor could always somehow obtain the probability of material error, and that the use of different sampling methods is not relevant to the analysis. See William R. Kinney, Jr. "A Decision-Theory Approach to the Sampling Problem in Auditing," <u>Journal of Accounting Research</u>, (Spring 1975), pp. 117-132; and William R. Kinney, Jr. "Decision Theory Aspects of Internal Control System Design/Compliance and Substantive Tests," <u>Journal of Accounting Research</u>, (Supplement 1975-Studies on Statistical Methodology in Auditing), pp. 14-34.

and the risk of a Type II error (the error of accepting a materially incorrect recorded amount) which is most frequently referred to as the combined risk; it may be appropriate to refine the internal control hypothesis further.[10] This is particularly true in light of the fact that a major distinction among strategies is how explicit they are in controlling the α risk. So, the following variants of the internal control hypothesis Ho are proposed:

> H_1: (weak form of the hypothesis): Internal control information can be used to reduce the statistical sample size of substantive tests without increasing the actual combined risk that arises as a result of using an audit statistical sampling strategy (relative to an "informationless" strategy).

> H_2: (strong form of the hypothesis): Internal control information can be used to reduce the statistical sample size of substantive tests without increasing either the actual combined risk or the actual α risk that arises as a result of using an audit strategy (relative to an "informationless" strategy).

Note that the hypotheses make no mention of the nominal or stated risk level associated with an audit strategy. A major result of the Neter-Loebbecke Study is that for typical accounting populations the nominal and actual risks can differ substantially when using classical statistical estimators. This implies that auditors cannot be certain how reliance on internal controls affects the actual sampling risks. Therefore, it appears that the hypothesis should be stated in terms of actual risks only. In terms of interpreting the research results, the strategies using internal control information should be keeping actual

---

[10] If an audit strategy consists of a substantive test only, then the combined risk reduces to the risk of Type II error as a result of the substantive test. The symbol β risk is frequently used to indicate a risk of Type II error for a substantive test.

risks at or below the actual risks associated with the informationless strategies. Thus the nominal risks will not be as relevant in evaluating strategies as they are in planning sample sizes employed by the strategies.[11]

The above formulation of the statistical form of the internal control hypothesis makes it possible to objectively measure the impact of internal control information on the performance of an audit strategy. That is, by examining the performance of various audit strategies that have been proposed or used in practice over a range of internal control conditions, it is possible to measure actual sampling risks associated with a given strategy and the average size of substantive test samples used by the strategy. With enough observations, it is possible to obtain important insight about the impact of internal control information by contrasting the performance of various strategies.

This framework provides a very simple way of assessing the relative performance of a sampling strategy: assuming actual sampling risks are not increased, performance is measured by the average substantive test sample size associated with a strategy over all internal control conditions. That is, that sampling strategy is preferred which for the

---

[11]This is because the auditor is really concerned with the actual sampling risks. It is an unfortunate fact of life that in many audit situations the nominal risk of a statistical estimator can differ significantly from the real risk. Therefore, the auditor is not always certain what his actual sampling risks are. This points out the need for more research on the performance of various statistical estimators under various audit conditions. This is an important area of audit research to which this study makes a contribution.

same amount of internal control information has a smaller average substantive sample size.[12] If the risks do increase with the use of internal control information, then evidence would be obtained that the internal control hypothesis does not hold (at least for the strategy considered). If the risks do not increase, then the research would provide estimates of the substantive sample size savings available as a result of internal control information. There is no need to assume costs of obtaining internal control information or making substantive tests. By using this approach and judiciously contrasting the performance of various possible strategies, it is possible to assess not only the impact of internal control information on a strategy but various other factors as well. This is attempted in the dissertation as reflected by the goals to be discussed next.

1.3 Major objectives of the dissertation

The primary goal of this dissertation is to test the validity of the often stated auditing assumption that information about internal control effectiveness can bring about reductions in subsequent audit work. In particular, the dissertation study is designed to examine whether such information may reduce the amount of substantive testing and maintain the reliability of the audit in a situation where statistical sampling is used (i.e., the statistical validity of the internal control hypothesis is tested).

---

[12]Actually, a stronger result holds by the way the strategies operate. For a given amount of internal control information, a sampling strategy using such information always has a sample size less than or equal to the related informationless strategy.

A closely related goal is to provide estimates of resultant efficiencies introduced by information on the internal control system. The impact of this information is measured by the reduction in the statistical sample size for substantive tests. The dissertation research is designed to indicate how different amounts of internal control information can affect the amount of substantive testing. The results should be useful in assessing the value of such information.

Another goal of the dissertation is to provide evidence for determining which of the selected methods (linkage rules) of linking information on internal control with substantive testing is most effective in terms of reducing sample size and maintaining the reliability of an audit. It is recognized that the value of information is a function of how that information is used. Thus it is necessary to also consider how the internal control information is used to affect substantive testing. Since no prior empirical research has been done on this issue, the proposed research will consider several of the linkage rules that appear to have the most support and test them for their relative effectiveness with different substantive testing methods.

Another goal of the dissertation is to provide estimates of the impact of uncertainties on the audit process brought on by subjective judgment. This information should be useful for assessing the value of introducing more objectivity (e.g., via formal internal control models or more audit training) to the audit process.

Finally, the dissertation provides evidence on the relative efficiencies of dollar-unit sampling (DUS), stratigies mean-per-unit (STMPU), and the recently developed Felix and Grimlund estimation technique which can be used within a Bayesian framework.[13] Although there has already been some extensive simulation work done concerning the relative performance of DUS and stratified MPU, the issue remains unsettled. The dissertation provides new data on the controversy by (a) using that variant of DUS which its advocates claim is superior to all others and (b) using a more realistic (externally valid) setting than those of earlier studies. A more realistic setting is accomplished by using the integrative approach specified in SAS No. 1 Sec 320.

In addition to these major goals, there are several lesser ones related to aspects of the specific models used and the type of data gathered in this dissertation, and to system reliability measurement. These goals are identified in the pertinent sections of the dissertation.

1.4 Research methodology

A computer simulation is used to achieve the goals of the dissertation. There are three reasons for this. First, computer simulation allows complete experimental control in defining the accounting environment and the sources of error in an accounting system. This is particularly important for accounting constructs because they frequently involve predictions in an uncertain environment. For example, in the real world the "true" value of an accounts receivable file may

---

[13] William L. Felix and Richard Grimlund, "A Sampling Model for Audit Tests of Composite Accounts," Journal of Accounting Research, (Spring 1977), pp. 23-40.

never be known and therefore the accuracy of an auditor's estimate of such value cannot be objectively ascertained. However, in a simulation, both the book values and the values that should have been recorded as far as the auditor is concerned (audit values) can be specified by the researcher. This allows for an objective measure of the accuracy of statistical estimates based on sample results--something which is not feasible in the real world.

Second, the relative performance of various combinations of methods and models represented by the audit strategies cannot be assessed analytically at the present time.[14]

Third, computer simulation has been the most common means of addressing similar issues in prior research and at this point it seems the best way of approaching the more diverse issues considered here.

The general approach of the simulation is to construct a population of records based on an actual accounts receivable file of a medium-size manufacturer. Each record contains a book value, audit value, and several fields indicating a processing trail. These fields provide information about compliance deviations in the system of internal controls for the records. A rule is used to determine the accuracy of the book values in the file by generating possible differences between book value and value that should have been recorded--audit value--of each record. The rule will be such that a classic audit situation is set up: total monetary errors go up with increases in

---

[14]Robert Kaplan discusses some of the issues relating to substantive tests in "Statistical Sampling in Auditing with Auxiliary Information Estimators," Journal of Accounting Research, (Autumn 1973), pp. 239-258.

compliance deviation rates and the internal control system provides evidence on the accuracy of the book values. Five internal control conditions are developed represented by five files of records. All five files have the same book values in common, but are differentiated by the compliance error rates and their associated monetary errors.

The five files will be "audited" by simulating different possible audit sampling strategies. Again, such an audit strategy is a combination of internal control information, linkage rule, and statistical substantive testing method that appears feasible in efficiently estimating the total audit value of each file.

A hypothesis testing approach is used in the simulation.[15] This simplifies the data gathering process for the audit strategies because no further assumptions need to be made about the simulated auditor's action if a file is rejected. Under this approach as many as three parameters (depending on the strategy used) need to be specified: α risk, β risk, and materiality level M. The simulation uses the following nominal values: $\alpha = .05$, materiality = M = 5% of total book value, and β is adjusted on the basis of internal controls so that combined risk = .05.[16] These parameter values appear to be the most widely

---

[15]A modified estimation approach will be used in assessing the compliance test results, however, in order to simulate estimates of the reliability of the internal control system. Justification for a hypothesis testing approach in substantive testing is provided by Robert K. Elliott and John G. Rogers, "Relating Statistical Sampling to Audit Objectives," Journal of Accountancy, (July 1972), pp. 46-55.

[16]Of course the actual sampling risks may differ significantly from these nominal values as pointed out earlier.

quoted at the present time.[17]

The use of the computer allows the simulation of many applications of an audit sampling strategy on an accounting file. Data can be collected about the results of using different audit strategies--for example data such as the average substantive test sample size, percentage of Type I errors, and percentage of Type II errors. This data will allow an analysis to be made on the impact of various factors on the performance of an audit strategy. The basic audit process as given in SAS No. 1 Sec. 320 (and paralleled by the audit strategies) will thus be analyzed via the simulation. In deciding which methods to use in a strategy, an attempt has been made to consider those methods which have the most support in the literature. With enough trials of the application of each of the strategies on each of the five files, a rather clear picture should emerge about the impact of different combinations of methods and amounts of internal control information.

1.5 Limitations and their possible significance

Up to this point there may have been some doubts generated about the generalizability of the simulation results given the many assumptions that need to be made about the accounting environment and the audit strategies simulated. This is a major problem associated with any simulation. The limited number of cases that can be considered in the analysis causes the results to be situation specific. The importance of the study thus rests considerably on the external validity and

[17]Support for this assertion is provided in chapter four of the dissertation.

representation of the accounting environments used and the audit strategies assumed.

In order to limit the scope of the dissertation of a manageable extent, it was decided to restrict the analysis to one basis population of book values. The accounting environments (represented by the five files) are based on an actual book value population encountered by auditors and it represents one of the more difficult populations (because of the high skewness and type of error pattern) for using statistical estimation. An even higher skewed population could have been used but it appears such environments are less typical of those encountered in actual audit situations.[18] Thus there is a certain amount of subjectivity involved in deciding on a single book value distribution. However, it is hoped that there is at least agreement that this population, because of the nature of its origins (i.e, accounts receivable, medium sized firm), is important per se and must be considered in evaluating the performance of an audit strategy.

Another major problem of using the simulation methodology (and of all models for that matter) is the necessity to make certain assumptions about the process being simulated. This is a particularly grave problem in this study because of the lack of empirical research on error rates and, more critically, the distribution of dollar errors in accounting populations. Hence, somewhat arbitrary assumptions about

[18]For example, three of the four Neter-Loebbecke populations have equal or less skewness than the one used here.

the effects of compliance error rates and dollar error generating processes must be made. These assumptions are only somewhat arbitrary because, fortunately, some limited guidance does exist from practitioners as to the most common types of error patterns. Pains have been taken to incorporate this information in developing the five accounting files.[19] In addition, the files reflect the classic audit situation wherein the internal control system provides evidence for the accuracy of the book values. It is assumed these factors result in a series of reasonably representative accounting environments.

However, perhaps the most important feature of the simulated accounting system may prove to be that it is a closed system and hence all sources of error arise from breakdowns in internal control. This means internal control information should have its maximum impact in such a system, and so if the internal control hypothesis were found to not hold under this most favorable circumstance, prima facie evidence would be provided that it wouldn't hold in <u>any</u> situation. This arises because the hypothesis is stated with such generality; it is not conditional on any particular accounting environment or audit method. Thus in testing the hypothesis, any reasonably representative accounting system can be used. Of course, the more such systems that are tested, the more evidence would be made available on the validity of the hypothesis, but this is true of all empirical research and the scientific

---

[19] It is interesting to note that these error patterns also turn out to cause the most problems for statistical estimators. See Paul John Beck, "A Critical Analysis of the Regression Estimator in Audit Sampling," (Ph.D. dissertation, University of Texas, 1977) chapter six.

method in general.

It is felt that the limited number of strategies considered is not as serious as the limited number of environments considered. This is because a review of the literature indicates it is possible to identify those strategies having the most support. Since at the present time no analytical methods have yet been devised for finding a single best audit strategy, it is felt desirable to at least assess the relative performance of those strategies having the most authoritative support under various relevant environmental conditions. The best present tool for doing this, if the assumptions can be agreed upon, is by computer simulation.

This research is distinguished from earlier similar research by the fact that three parameters--α, combined risk, and materiality--have been specified in advance and one might argue that this may reduce the generality of the research effort. However, this appearance of loss of generality is deceptive for it turns out that the assumptions and interpretation of the earlier results are themselves open to some dispute. For example, Kaplan used an indirect measure of the general tendency of an estimator to make a Type II error by measuring the correlation between the estimates of the standard error and estimates of the population mean or total.[20] However, more recent research by Beck found that such correlation statistics do not provide a good measure of the actual Type II risk in typical accounting environments.[21] This appears to be

[20]Kaplan, pp. 239-258.

[21]Beck, pp. 192-194 and pp. 179.

due to the fact that such risk is affected not only by the distribution of book values but by the particular error pattern as well. Hence a correlation measure alone is not sufficient for predicting the actual Type II risk.

For another example, the approach used by Neter and Loebbecke measures the reliability of an estimator (that is, the frequency with which the population audit value is contained within the confidence interval) and collects data on the sampling distribution of the estimator. This apparently more general method (because it does not make assumptions about what is materiality) has several disadvantages as far as audit applications are concerned. First, their approach does not directly measure the actual α and β risk associated with a sampling method. Since the seriousness of these risks is considered to be significantly different in an audit context, it appears that an direct approach which does directly measure both of these risks is more relevant to auditors. This is reflected by the way the internal control hypothesis has been formulated. One is hard pressed to find an equivalent formulation of the internal control hypothesis not considering the actual sampling risks. In fact, a review of audit training manuals indicates most audit firms as a matter of firm policy use specified α and combined risks in planning their sample sizes. Thus it should not be inappropriate to use the same approach in simulating the sampling strategies.

A second disadvantage of the Neter and Loebbecke approach is that it does not reflect the typical audit situation wherein the sample sizes tend to vary with the amount of error in a population as a result of internal control information. Thus it is felt that a more valid performance of a sampling method is obtained where internal control information can influence the amount of sampling through linkage rules. This also requires specification of α, combined risk, and M.

Finally, it should be noted that what initially appears to be a more general approach requires some rather arbitrary assumptions of its own to be made--namely, the constant sample size to use and the nominal reliability level for constructing a confidence interval. The difference between the two approaches thus reduces to the kinds of parameters that should be specified in evaluating a sampling method. The researcher essentially argues that the more relevant approach is the one followed by most audit firms using statistical sampling in practice and implied by audit theory. For these reasons the direct approach is used in the dissertation and, therefore, α, combined risk, and materiality values are specified in advance.

Another basis for questioning the approach used in this dissertation might be the apparent failure to identify a loss function. This is not a completely correct characterization of the study. By specifying the α, combined risk, and M values, a constraint is effectively put on the form of the loss function. What a more specific representation of the loss function should be is open to question; in fact, at present there is not even agreement as to whose loss function should be used in the evaluation of audit samples.[22] Therefore, it appears

---

[22]For example see Robert P. Magee, "Discussion of Auditors' Loss Functions Implicit in Consumption Investment Models," Journal of Accounting Research, (Supplement 1975--Studies on Statistical Methodology in Auditing), pp. 121-123.

preferable to take the approach of specifying particular α, combined risk, and M values, assuming there is a greater likelihood that general agreement can be reached on the validity of such values than on the form of the loss function. That this is reasonably possible is supported in chapter four of the dissertation.

Another possible limitation of the study is that nonsampling error is not more completely considered in the performance of sampling strategies. Some have pointed out that this is likely to be the biggest component of error in the audit sampling decision problem. Unfortunately, there is little bheavioral data on which one can base a simulation introducing such errors. In this dissertation a pioneering attempt is made to assess the impact of introducing such errors by making a rather arbitrary error generating process assumption. Nevertheless, it is hoped that this at least provides a start in addressing the importance of the nonsampling risk issue.

Yet another limitation is that cost is not directly incorporated in the analysis. It is hoped that the substantive test sample size provides a good surrogate for cost. Certainly there are drawbacks with this approach; for example, no distinction is made between drawing a dollar unit sample or a record unit sample--a distinction which may have considerable cost consequences. Similarly, no assumptions are made about auditing the sample or obtaining internal control information pertinent to the sample. On the other hand, this apparent limitation at the same time increases the generality of the results because any practitioner or researcher can evaluate a sampling strategy based on the results of this study by using his own cost structure. Nevertheless, in principle, a complete audit decision model should incorporate the relevant costs.

The question of the importance of this research reduces to whether there is a real need for obtaining a reasonable assessment of the relative performances of valid audit sampling strategies in sufficiently realistic accounting environments. Considering that presently there is so little known about the validity and robustness of the internal control hypothesis (with accompanying implications for linkage rules, substantive testing methods, and value of internal control information); the researcher believes that the answer is, "yes".

1.6 Outline of the dissertation

In chapter two, a review of prior similar research is given. First there is a review of some of the internal control models that have been proposed and the theoretical justification for obtaining internal control information. Second, there is a review of statistical sampling research in auditing. Third, some pertinent behavioral studies in auditing are reviewed. Finally the chapter concludes with a discussion of the relationship of the dissertation research to prior research.

Chapter three describes the accounting environments that are simulated. All assumptions about what constitutes a reasonably representative accounting environment are stated and reasons therefore given. Appendix I shows that this is a sufficiently rich environment for testing the general statistical validity of the internal control hypothesis.

Chapter four discusses the audit sampling strategies that are simulated. The strategies are described in detail and reasons for choosing this particular set of strategies are given. The discussion also covers issues pertaining to measurement of overall system reliability based on a set of compliance tests.

Chapter five describes the empirical findings. The performances of the strategies are compared and data pertaining to various aspects of the performances are presented. The data is then used to address the issues represented by the objectives stated in section 1.3 of chapter one.

Chapter six summarizes the overall findings of the research. The ramifications of the research results for auditing theory and practice are then discussed. The dissertation concludes with a discussion of potential extensions of research which are indicated.

Appendix II shows that assuming the compliance error rates in an internal control system are equal results in a conservative estimate of system reliability. This result provides a justification for using equal error rates in the simulated accounting environments.

Appendix III is a glossary of the key terms used in this dissertation.

Appendix IV is a presentation of the Mann method for computing lower confidence bounds on series system reliability that is used in the simulation.

Appendix V is a presentation of the Felix-Grimlund model formulas used in the simulation.

Appendix VI provides the statistical decision rule under the negative approach using stratified mean-per-unit estimation.

Appendix VII provides an intuitive explanation for the growth of actual α risk as the immaterial error size increases, and graphs such growth for the substantive test methods used in the simulation. It also develops a sample size formula for controlling α risk for any immaterial amount of error when using stratified mean-per-unit estimation.

CHAPTER TWO

Review of Prior Related Research

2.1 Preface

This chapter provides an overview of the pertinent auditing literature relating to the dissertation research. In many cases the proposed methods stand in isolation because no attempt has been made to relate these methods to other aspects of the audit. In effect, little has been done to make the proposals suitable for practice and, therefore, the potential benefits remain largely unknown. Thus a major deficiency of much of the prior audit research is the failure to provide guidance to audit practice. And this in part may be due to the fact that auditing theory itself is not rich enough in specifying assumptions and yielding testable hypotheses.

This chapter organizes the literature from the perspective of an internal control theory within an auditing framework. Having identified the hypothetical value of internal control information to the auditor, the various normative internal control models are then reviewed. The next section reviews the results of behavioral research which provides evidence on the behavioral validity of the internal control hypothesis. Then the literature on statistical sampling in auditing is reviewed and related to the internal control hypothesis. The final section summarizes the ramifications of the literature for the validity of the internal control hypothesis and concludes with an outline of the research steps that are indicated.

2.2 Internal control within the framework of auditing theory

Internal controls can be viewed from the perspective of either (1) the independent auditor; or (2) the managerial accountant, internal auditor, or systems analyst. The main difference between the two views is in the number of controllable factors that are considered worthwhile for monitoring and evaluation. Generally, the independent (external) auditor is primarily concerned with only a subset of the internal controls considered by the others. However, if it is believed that the scope of the audit function will eventually expand to include such things as management or operational audits, then it becomes apparent that the two control concepts are similar.

It should be noted at this point that the internal control concept does cross into other functional areas of accounting (as indicated by the two views discussed in the preceding paragraph). Hence it is conceivable and perhaps even desirable that an internal control theory be developed independent of auditing theory. Therefore, although the dissertation restricts itself to analyzing the implications of internal controls within the context of auditing theory only; one should remember that the results may have import reaching beyond the immediate auditing setting. A broad internal control definition is the following:

> "Internal control comprises the plan of organization and all of the coordinate methods and measures addopted within a business to safeguard its assets, check the accuracy and reliability of its accounting data, promote operational efficiency, and encourage adherence to prescribed managerial policies. This definition

> possibly is broader than the meaning sometimes attributed to the term. It recognizes that a "system" of internal control extends beyond those matters which relate directly to the functions of the accounting and financial departments.[1]

This definition appears capable of encompassing many factors over which management has control and thus will be considered sufficient for the purposes of present auditing theory.[2] The more recent Statement on Auditing Standards No. 1 (SAS No. 1) distinguishes two facets of internal control: administrative controls and accounting controls.

> "Administrative control includes, but is not limited to, the plan of organization and the procedures and records that are concerned with the decision processes leading to management's authorization of transactions. Such authorization is a management function directly associated with the responsibility for achieving the objectives of the organization and is the starting point for establishing accounting control of transactions.
>
> Accounting control comprises the plan of organization and the procedures and records that are concerned with the safeguarding of assets and the reliability of financial records and consequently are designed to provide reasonable assurance that:
>
> a. Transactions are executed in accordance with management's general or specific authorization.
>
> b. Transactions are recorded as necessary (1) to permit preparation of financial statements in conformity with generally accepted accounting principles or any other criteria applicable to such statements and (2) to maintain accountability for assets.

---

[1] American Institute of Certified Public Accountants (AICPA), Statement on Auditing Standards No. 1, (New York: AICPA, 1973) Sec 320.09.

[2] The researcher is aware that broader definitions exist. For example, that adopted by internal auditors in the Statement of Responsibilities of the Internal Auditor would include checks on the accuracy and reliability of all data, not just accounting data.

c. Access to assets is permitted only in accordance with management's authorization.

d. The recorded accountability for assets is compared with the existing assets at reasonable intervals and appropriate action is taken with respect to any differences."[3]

This dichotomization lends itself to a comparison of the internal controls the independent auditor is primarily concerned with, accounting controls, with controls of primary interest to management, administrative as well as accounting controls. This restriction of scope to internal accounting controls reduces the number of controls the auditor needs to evaluate, simplifies the development of internal control models, and perhaps most importantly, permits a reduction of the legal responsibility on the auditor. It is this subset of internal controls which has been the subject of recent official pronouncements on internal control issues and also the object of modeling efforts in this area.

Although it may be desirable to use a narrower definition of internal control to reduce legal problems for the auditor, the legal approach does not necessarily provide a sound basis for theoretical developments. Legal requirements can change: witness, for example, court cases deciding what is material and fair in financial statement presentation.[4] Thus, a theory based on auditor's liability is subject to the vicissitudes of every new court opinion.

---

[3]AICPA, Auditing Standards, Sec. 320.10.

[4]For example, see Larry Rittenberg and Bradley Schweiger, "Auditor Reporting Responsibilities in a Changing Society," Wisconsin Working Paper 7-76-26, Graduate School of Business, University of Wisconsin-Madison, 1976, pp. 11-14.

A reasonable approach to building an auditing theory of internal control is to consider the use of such information in an audit context. Internal control information affects three major audit decision and reporting functions:

> "(1) To determine whether an audit is possible
> (2) To determine the scope of the audit
> (3) To make recommendations to management."[5]

Of these, the second is the most important justification for obtaining internal control information in the normal audit:

> "The purpose of the auditor's study and evaluation of internal control as expressed in the auditing standard quoted in paragraph .01, is to establish a basis for reliance thereon in determining the nature, extent, and timing of audit tests to be applied in his examination of the financial statements"[6]

This justification is reflected in just about every auditing text, auditing theory, and in the second auditing standard of fieldwork. It is the reason for the existence of the internal control hypothesis.

The rationale of the internal control hypothesis is that the auditor "cannot determine how much work to do or what kind of work to do until he has become familiar with the strong and the weak points of the internal control system which protects the enterprise resources and provides the data on which he is asked to present an opinion."[7] Thus the internal control evaluation is a necessary first step to any audit engagement.

---

[5]Taken from Alvin A. Arens and James K. Loebbecke, Auditing, An Integrated Approach (Englewood Cliffs, New Jersey: Prentice Hall, Inc., 1976), p. 160.

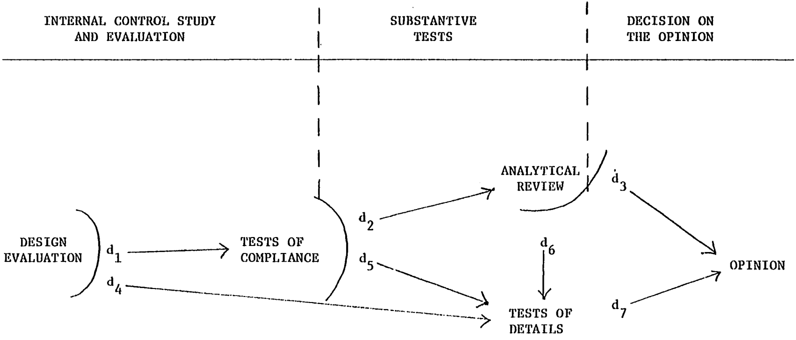
[6]AICPA, Auditing Standards, Sec. 320.06.

[7]R.K. Mautz and Hussein A. Sharaf, The Philosophy of Auditing (Sarasota, Florida, American Accounting Association, 1961), p. 142.

The audit literature tends to stress the "how much work" aspect of internal control information usefulness. There is a strong implication that work efficiencies are introduced by internal control analysis, that without such an evaluation the auditor would not use the available time and personnel as effectively as possible. It is this assumption that the internal control hypothesis articulates and the research addresses. Hence the hypothesis does not deal with other aspects of the usefulness of internal control information to the auditor such as for determining the kind of work to do or serving as a vehicle for making recommendations to management.

To better explain how internal controls can affect the amount of audit work the audit process as specified in SAS No. 1, Sec 320 will now be reviewed. The financial statement opinion audit contains three basic stages: (1) internal control study and evaluation, (2) performance and evaluation of substantive tests, and (3) a decision on the form of the opinion (see figure 1). The internal control study and evaluation stage can in turn be subdivided into design evaluation and tests of compliance. Similarly, the substantive test stage can be subdivided into analytical review and tests of details. Design evaluation consists of a review and preliminary evaluation of the internal control system design. Compliance testing consists of tests (judgmental or statistical) to provide assurance that internal controls are in use and operating. Analytical review comprises the analysis of significant ratios and trends and resulting investigations of unusual relationships and questionable items. Tests of details of transactions and balances

Fig. 1: Routes to an Opinion on an Account Under SAS No. 1, Section 320



d_i, $i = 1, 7$ represent decisions that need to be made on the route to an opinion

---

SOURCE: William R. Kinney, Jr., "A Decision-Theory Approach to the Sampling Problem in Auditing," Journal of Accounting Research, Spring 1975, pp. 117-132, figure 1.

consist of making tests of the system for the purpose of assessing the dollar accuracy of the records. Of these various audit procedures, the substantive tests are the most basic and important because they represent the primary form of evidence for an auditor's opinion and they are required for inventories and receivables by auditing standards. The first two stages are thus characterized by the kind of evidence gathered by the auditor to support his opinion on the fairness of financial statement presentation.

The internal control hypothesis of auditing theory essentially states that the more reliable the system of internal controls the less extensive the substantive tests the auditor needs to conduct. As indicated earlier this reflects the intuitively appealing notion (for management as well as auditors) that a good internal control system should require less extensive analysis and testing than a bad one. This is operationalized in statistical samples by varying reliability levels inversely with reliance on internal controls as indicated in chapter one.

There are presently many methods for evaluating internal accounting controls ranging from using only the auditor's professional judgment to formal mathematical models. The methods vary in sophistication and accuracy, (and, presumable, cost) and thus in order to determine if a particular model is cost justified, it is necessary to assess how the quality of information about the internal accounting controls introduces savings, if any, to the substantive testing stage of the audit. This assessment is possible only in situations where statistical sampling is used for substantive testing because only then does there exist an objective basis (as defined by statistical theory) for testing the internal control hypothesis. In addition, since auditing is tending to rely more on statistical techniques, the way internal control affects the use of these techniques is becoming a separate issue of importance in its own right as well as perhaps providing indirect evidence of the validity of the hypothesis when judgmental sampling is used.

Conceptual framework of audit linkage rules. The restriction of the analysis of the internal control hypothesis to a situation where statistical sampling is used retains many of the unresolved issues in auditing theory. To determine the value of information generated by the formal mathematical models, it is necessary to determine how the extent of substantive testing is affected by the amount of internal control information. In fact the internal control hypothesis is inseparable from the linkage issue. There have been several linkage methods proposed in the literature and perhaps a greater variety of methods are actually used in practice. These are discussed after reviewing the formula proposed in the auditing standards to illustrate the quasi-Bayesian conceptual framework of the linkage relationship.

According to auditing standards:

> .34 The auditor's judgment concerning the reliance to be assigned to internal accounting control and other relevant factors should determine the reliability level to be used for substantive tests. Such reliability should be set so that the combination of it and the subjective reliance on internal accounting control and other relevant factors will provide a combined reliability level conceptually equal to that which would be used in the circumstances described in paragraph

.32 [a situation where no reliance is placed on internal control]. Thus the reliability level for substantive tests for particular classes of transactions or balances is not an independent or isolated decision; it is a direct consequence of the auditor's evaluation of internal control, and cannot be construed properly out of this context.

.35 The concept expressed in paragraph .32 can be applied by use of the following formula:

$$S = 1 - \frac{(1-R)}{(1-C)}$$

where S = Reliability level for substantive tests.

R = Combined reliability level desired (e.g., 95 percent as illustrated in paragraph .32)

C = Reliance assigned to internal accounting control and other relevant factors.[8]

This formula is interesting because it reveals a lot about the auditing profession's conceptualization of the audit process. If the formula is rewritten into the equivalent form (1-R)=(1-S) (1-C), the multiplicative relationship of the risks is emphasized. SAS No. 1, Sec 320 B described the nature of this relationship as follows:

> The ultimate risk against which the auditor and those who rely on his opinion require reasonable protection is a combination of two separate risks. The first of these is that material errors will occur in the accounting process by which the financial statements are developed. The second is that any material errors that occur will not be detected in the auditor's examination.
>
> The auditor relies on internal control to reduce the first risk and on his test of details and his other auditing procedures to reduce the second. The combined risk of both of the related adverse events occurring jointly is the product of the respective individual risks,

---

[8]AICPA, Auditing Standards, Sec. 320 B .34 and .35 where reliability in the standards is defined to be "The proportion of such ranges (intervals) from all possible similar samples of the same size that would include the actual population value". Sec. 320A.03. This reliability in statistical decision terms is the complement of the β risk associated with the statistical test (See Sec. 320B.30 or Don M.

and the combined reliability is the complement of such combined risk. [This last relationship is best brought out by using yet another equivalent formulation: R=1-(1-S) (1-C)][9]

There are some interesting assumptions buried in the above formulas and interpretations. For one thing, it appears the AICPA's Auditing Standards Executive Committee (henceforth, the committee) is assuming the two risks are independent since the combined risk is the product of the two. This is so because any dependence between the two risks would not guarantee such a functional form for the ultimate risk. Another assumption the committee made was to ignore the risk of making a Type I error (the risk of deciding there may be a material error when the net amount of monetary error is immaterial) by only considering the risk (1-S) of not detecting material errors during the substantive testing stage of the audit. Thus the risk of doing too much substantive testing was apparently assumed to be not nearly as important as the Type II risk.

Possibly the most revealing assumption made in using the formula is that the audit process reliability can be represented as the reliability of an independent parallel system in which the operation of a particular component (i.e., stage in the audit process) does not depend on the operation of the other components.[10] What this means is

---

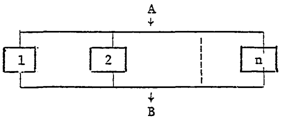
Roberts, "A Statistical Interpretation of SAP No. 54," Journal of Accountancy, (March, 1974) pp. 47-53.

[9]AICPA, Auditing Standards, Sec 320B.29

[10]A parallel system of n components is one in which the components are as follows:

that the reliability of the audit--combined reliability--is a function of the reliability of the internal control system and the reliability of the substantive tests in detecting a material error. Hence an auditor's opinion would be in error only if <u>both</u> the internal control system and the auditor's substantive test failed to detect a material error. The first stage of the audit process determines how much reliance should be placed on the internal control system. The second stage determines the amount of reliance to put on substantive testing. The formula thus provides a conceptual framework for linking internal control information with substantive testing.

However, the formula does not seem to be literally used in audit practice. The chief difficulty appears to be in operationalizing an objective measure of C, reliance on internal control and other audit procedures. This is done for the most part indirectly via subjective interpretation of compliance test results and other audit evidence.



Thus if one imagines a process which must travel from point A to point B, the process will be successfully completed as long as one of the n components is operating satisfactorily.

In contrast, a serial system of n elements is organized as follows:

$$A \longrightarrow \boxed{1} \rightarrow \boxed{2} \longrightarrow \text{---} \longrightarrow \boxed{n} \longrightarrow B$$

This kind of system fails whenever any one of the n components fails.

The reliability of a system can be defined to be the probability

This process is described in the auditing standards as follows: "The samples should be evaluated in terms of frequency and nature of the deviations from any procedures the auditor considers essential to his preliminary evaluation of internal control, and that their influence on his final evaluation of internal control should be based on his judgment as to the effect of such deviations on the risk of material errors in the financial statements."[11] Thus it is necessary for the auditor to convert internal control information, which frequently consists of error frequency data, to an assessment of the C value for use with the SAS No. 1 Sec. 320 linkage rule.

Since this involves subjective estimates for interpreting the error rates and analyzing the internal control system (e.g., some internal control procedures depend primarily on segregation of duties and leave no audit trail of documentary evidence of compliance), several classes of techniques have been developed in practice to operationalize the SAS No. 1 linkage rule.

One way of classifying these techniques is on the basis of the selection method for compliance testing and the degree of flexibility in interpreting the test results. The following figure summarizes the alternatives that are available under these two bases.

---

of successful processing. It should be noted for clarification that the reliability presently being discussed is that associated with the audit process; later, the reliability of an internal control system within an organization is discussed.

[11]AICPA, Auditing Standards, Sec. 320A.22

Fig. 2. Linkage Rule Classification

	Degree of Flexibility in Interpretation	
Selection Methods	None (Rigid)	Flexible
Value-oriented (probabilities proportional to size of account—dollar unit sampling)	Alternative 1	Alternative 3
Neutral (Simple random sampling)	Alternative 2	Alternative 4

The selection method refers to whether simple random sampling is used for compliance testing or value oriented selection where the probability of selection is proportional to the size of the book value of the account or transaction. Note that under value oriented selection the compliance error rate is biased toward measuring compliance for the larger dollar value items. This has been justified on the ground that there is a strong link between the size of transaction processed and the dollar error associated with the transaction.[12]

Rigid interpretation refers to those rules which make an assumption about the relationship between compliance error rates and monetary

[12] In fact this is an implied assumption for the standard procedure in auditing of making a 100% examination of very large items. See for example Rodney J. Anderson, The External Audit 1, Concept and Techniques (Pittman Publishing, Toronto, 1977), p. 344.

error rates which is assumed to hold for all accounting environments. Generally, these methods are conservative in that reliance is not planned unless the compliance error rates are extremely low no matter what the nature of the system facing the auditor. Rigid interpretation is one viable response available to auditors in dealing with highly uncertain environments. However, it may result in excessive substantive testing (given that the auditor does compliance testing). Nevertheless, rigid interpretation appears to be a fairly commonly used in practice with both selection methods. This is particularly true of value oriented selection where the most common such assumptions are that either (1) each compliance deviation results in a complete 100% overstatement of the associated transaction or, the less conservative assumption, (2) one in three compliance deviation results in a complete 100% overstatement of the associated transaction.[13] An example of a rigid interpretation is the following example given by Robertson:

"One major public accounting firm has expressed the following policy on acceptable compliance rate of error.

Auditor's Judgement of Required Degree of Compliance	Recommended Range for the Acceptable Upper Limit of Error
High	2% to 4%
Intermediate	5% to 8%
Low	9% to 12%

---

[13] Assumption one is described in Haskins & Sells, Audit Sampling, A Programmed Instruction Course, (Haskins & Sells 1970) frame 3-97. Assumption two is described in Clarkeson, Gordon & Co., Audit Testing Course, (Clarkeson, Gordon, & Co. 1975) Chapter XX, in particular, p. 119.

These evaluations of internal control are then converted to β values for use in planning substantive test sample sizes. For example, if the upper error limit is .02, β may be set to .5; if the upper error limit is .03, β may be set to .33; if the upper error limit is .04, β may be set to .25; if the upper error limit is .05, β may be set to .20; and if the upper error limit is greater than .06, β may be set to .05 (assuming combined risk is planned at .05).[14]

Flexible interpretation rules, on the other hand, are defined to be those which are tailor made for the particular accounting environment facing the auditor. That is, an interpretation of compliance test error rates is made trying to establish the particular relationship between procedural deviations detected in the compliance tests and the monetary errors which might exist in the accounting records of the specific firm being audited. It is thus more in line with the Bayesian flavor of the SAS No. 1 linkage. Needless to say, this is a difficult task fraught with potential judgmental errors, and perhaps this difficulty explains the popularity of the rigid approaches which typically represent an audit firm policy decision and reflect years of experience in working with many systems.[15]

---

[14]Jack C. Robertson, Auditing, (Business Publications, Inc., Dallas, Texas, 1976) p. 263 and p. 369. Also AICPA, Auditing Standards, Sec. 320B.22 has an example of what could either be a flexible (specific to the audit situation) or rigid (specific to the audit firm) interpretation.

[15]The rigid interpretations are essentially consistent with the philosophy of concentrating on limiting the risk of Type II errors at all costs. This is so because the risk of unwarranted reliance on internal controls is practically eliminated by the conservatism of the

Both interpretations link the frequency of compliance errors to the risk of material monetary errors and are thus in conformity with auditing standards. On the surface these methods may look far removed from the linkage formula of SAS No. 1, but they all represent different ways of operationalizing this rule given real world constraints. However, since these methods do not explicitly assign a C vlaue and use the SAS No. 1 linkage, there remains the empirical question of how they compare in performance and validity.

The simulation will include examples of alternatives (1), (3) and (4) of the linkage rules. Since flexible interpretation can always increase the value of internal control information (assuming no subjective errors), these are the ones stressed in the research. The least conservative example of alternative (1) is also included because one version of this approach has been given considerable support in the Canadian audit literature and the simulation relationships are closely related to these assumptions. The particular linkage rules used in the simulation are described in more detail in chapter four.

---

techniques (conservative in the sense not relying on internal controls to reduce substantive testing when some reliance may be possible).

Rigid rules represent an understandable response to the complexities of auditing in the real world. What is more they are one way of promoting standardization and consistency of audit practice within an audit firm. This may be a particularly important factor in explaining their popularity considering the high turnover of personnel many auditing firms experience. Thus rigid rules may also represent a response to controlling the quality of audit practice.

2.3 Review of internal control models of auditing and related theoretical bases

Review of the models

All of the models discussed in this subsection suffer from the common limitation that they are suitable for modeling only those controls which have a dichotomous character. That is, the models can capture the impact of controls only if it is readily determinable when an element is in control or out of control--this is typically the case for internal accounting controls. However, administrative controls can frequently take on an infinite range of values and it is less clear when a given decision is "correct" or "incorrect". Therefore, the models are not as suitable for modeling internal controls (as defined on pp. 27-28) as they are for modeling internal accounting controls only.

This is not to say the models are useless. Evaluation of internal accounting controls is a major practical problem in its own right which auditors must address constantly. However, since the models are poorly suited for administrative controls, the point is that they would not be as useful for more comprehensive audits, such as management or operational audits, toward which the profession is evolving.

Having recognized this common limitation, the point will not be mentioned further because the dissertation is concerned with evaluating internal control information within the more narrow framework of determining the extent of substantive tests. Henceforth, the term internal control is used synonymously with internal accounting control unless specifically indicated otherwise.

Models of internal control systems can be divided into two groups, mathematical and nonmathematical models. The nonmathematical models of internal control systems have usually consisted of written and/or chart descriptions of organizational relationships. Organizational behavior theories have sometimes been brought into the analysis to provide a more formal approach to the assessment of internal controls than is usually done by the practitioner. Variations of nonmathematical models such as flowcharts, questionnaires, and narratives are the most commonly used models of internal control systems in practice. However, although they are useful for listing or summarizing internal control characteristics, they are not as useful for objectively integrating this information and reaching a conclusion about the performance of the overall system. The integration must be done judgmentally.[16] Examples of nonmathematical internal control models in the accounting literature include the positional analysis models of Mautz and Mini, Swieringa and Carmichael, and the behavioral model of Carmichael.[17]

---

[16] Objectivity in this dissertation is intended to have the same meaning that Ijiri and Jaedicke attached to the term. That is it refers to the degree of unanimity or variability in measuring an object. An objective measure or assessment has less variability or more unanimity than a subjective assessment. See Yuji Ijiri and R. K. Jaedicke, "Reliability and Objectivity of Accounting Measurements," Accounting Review, July 1966, pp. 474-483.

[17] R.K. Mautz and Donald Mini, "Internal Control Evaluation and Audit Program Modification," Accounting Review, April, 1966, pp. 283-291; Robert Swieringa and D.R. Carmichael, "A Positional Analysis of Internal Control," Journal of Accountancy, Feb. 1971, pp. 34-43; D. R. Carmichael, "Behavioral Hypothesis of Internal Control, Accounting Review, April, 1970, pp. 235-245.

Mathematical models, on the other hand, have a higher potential for more objectively evaluating internal control systems. Several such models have been proposed in recent years paralleling the general trend in auditing and accounting toward the development of more objective techniques such as statistical sampling. Since only the mathematical model performance can be evaluated in an objective manner, these will be the only ones discussed in more detail.

Yu and Neter appear to be the first to have attempted to use a completely developed mathematical model for internal control evaluation purposes.[18] They applied Markovian theory to the problem by conceiving the internal control system as a finite stochastic process to which the Markovian property applies.[19] At any point in the processing cycle, an operating probability matrix is used to represent both the state of a record as input and the probability of changes in the "error" state resulting from processing. The output of the model is a vector of probabilities that various states of a record will occur.

Although this appears to be the first objective internal control evaluation model to be developed, there are several drawbacks to it

---

[18]Seongjae Yu and John Neter, A Stochastic Model of the Internal Control System," Journal of Accounting Research, Autumn 1973, pp. 273-295.

[19]A Markovian property for a process can be defined as one in which the conditional probability of any future "event" depends upon only the present state of the system.

which might explain perhaps why this model has not yet been adopted in practice: (1) the need to specify a separate transition matrix for each control point thus making the model cumbersome to apply in practice; (2) the potentially huge data requirements necessary to specify the numerous transition matrices which one would normally need to use to model a typical control system; and, perhaps most seriously, (3) the necessity to assume perfect knowledge of the probabilities associated with each state.

Another model for internal control evaluation was that introduced by Cushing.[20] He used reliability theory from engineering to analyze the overall reliability of the internal control system.[21] The model presents formulas to compute the overall reliability of a system as a function of the types of control components used, and the types of interrelationships between these control components. This model is very similar to the Neter-Yu model except that the output restricts itself to an estimate of the probability of a single state occurring, instead of a vector of probabilities for a set of states. In effect the Cushing model uses reliability theory to estimate a single point

---

[20] Barry E. Chushing, "A Mathematical Approach to the Analysis and Design of Internal Control Systems," Accounting Review, January 1974, pp. 32-45.

[21] The following general definition of system reliability was first given in footnote 10: the probability of successful processing. More specific definitions can be developed for auditing use and will be discussed later. However, the most frequent meaning attached to internal control system reliability in an audit context (and the one implied by Cushing) is the following: the probability of processing a transaction or record without producing a monetary error. A more general auditing definition is given on p. 54.

from the set estimated by the Neter-Yu model.[22] This perhaps makes the model more suitable for the auditor because it allows him to focus on estimating that parameter which is of most usefulness to him. However, the same basic criticisms apply as to the Neter-Yu model, the exception being that the model is less cumbersom because it concentrates on estimating the probability of only one state occurring (that usually being the probability of processing a transaction without a monetary error).

Bodnar more completely considered the problems of applying the reliability model to a system having a human element.[23] He concluded that a reliability modeling approach is feasible although major implementation difficulties may be encountered.

A recent dissertation by Stratton reported on a field study which involved the evaluation of an actual internal accounting control subsystem (an order entry system) using reliability theory.[24] The intent of the study was to demonstrate the feasibility of applying this tool to a real life accounting environment. Stratton felt he succeeded in showing this because he not only found evidence that the actual system

---

[22]Grimlund formally proved this. See Richard A. Grimlund, "A Framework for the Integration of Auditing Evidence," (Ph.D. dissertation, University of Washington, 1977) pp. 147-150.

[23]George Bodnar, "Reliability Modeling of Internal Control Systems," Accounting Review, October 1975, pp. 753-768.

[24]William O. Stratton, "Accounting Internal Control Systems: Their Reliability and Dichotomic Structure Functions," (Ph.D. dissertation, Claremont Graduate School, 1977).

satisfied the constant reliability assumption of the theory, but it was also possible to introduce substantial savings in analysis by focussing on a few critical processes. He concluded that the model showed considerable promise for use in objectively evaluating accounting internal control.

However, Stratton had a few technical problems which seriously threatens the accuracy of his analysis.[25] At some points he appeared to use ad hoc procedures. On the other hand this may reflect the limitations of the theory. In spite of these problems, the field study is a valuable contribution to the auditing literature because it illustrates the problems as well as the promise that exist in applying reliability theory to internal control system analysis.

Another internal control model is that by Burns who has proposed the use of a computer simulation to assist the auditor to evaluate an internal control system.[26] He has constructed two simulation models for an inventory system. A computer simulation model simulates the entire inventory system and produces as output the typical accounting reports obtained from such a system. These reports contain errors as some function of the probability of an error in the individual control points of the system. Errors are generated by the Monte Carlo method.

---

[25] Ibid., p. 140. It should also be noted that since Stratton did not know the true system reliability there is no way of measuring the accuracy of his calculations.

[26] David C. Burns, "Computer Simulation: A Tool for Testing the Effectiveness of Internal Control Auditing Methods," Proceedings of the Fifth Annual Midwest AIDS Conference, Vol 1, Minneapolis, MN, 1974, pp. D1-D4; and David C. Burns and James K. Loebbecke, "Internal Control Evaluation: How the Computer Can Help," Journal of Accountancy,

An audit simulation model is used to simulate the internal control system. This model gives an estimate of overall system reliability and the amount of dollar error produced by the system that the auditor can use to plan the remaining audit work. Undoubtedly, the accuracy of the audit simulation model is affected by the fact Burns created and defined the inventory system which is being modeled. Auditors frequently do not have this degree of intimacy with the system they are auditing.

Nevertheless, the model holds great promise for two important reasons: (1) the model allows the auditor to obtain a direct estimate of the relationship of compliance errors to the amount of dollar error for a given set of error probabilities--this allows the auditor to more readily estimate the C value in the SAS No. 1 Sec 320 formula; and (2) the computer methodology provides a mechanism for exploring the implications of variations of auditor's judgment of his uncertainty. Thus the auditor does not have to know the system relationships with perfect accuracy because by means of extensive sensitivity analysis he can translate his uncertainty into its ultimate dollar impact. The presumption is that if sensitivity analysis does not indicate there is a strong impact on dollar error then the auditor's uncertainty about the parameter values is not important and the system can be relied upon.

Another approach which like the simulation approach provides the auditor with an estimate of the possible dollar error in the financial statements, but which does this analytically, is the model developed by

---

August 1975, pp. 60-70.

Grimlund. In reality there are several models which Grimlund introduced--all of them attempting to achieve the same purpose of integrating audit evidence to determine implications on account balances, but differing in their approaches. However, only one of these, the beta-normal model, was completely developed in Grimlund's work and therefore this is the one that is considered here.

Grimlund's beta-normal model represents the most ambitious attempt yet to model the various sources of audit evidence in a comprehensive manner. It integrates both the indirect evidence of internal control evaluation (essentially via the beta distribution) as well as the more direct evidence available from substantive tests (essentially via the normal distribution). It is thus the most complete analytical model of the audit process available. It goes significantly beyond the previous non-simulation models in that its output consists of an estimate of the amount of dollar error the system could produce and therefore the model is of more potnetial usefulness to the auditor.[27]

Unfortunately, the model's usefulness remains in the realm of the potential only because it still awaits implementation on a real world system. Its practicality still needs proof. There is also no

---

[27]An auditor's basic objective in the usual audit is to obtain sufficient evidence on which to base an opinion on the fairness of the presentation of the financial statements. Internal control information provides less direct evidence than substantive tests on financial statement accuracy. Implied in this setting is that the auditor should establish some relationship between his assessment of internal controls and the accuracy of the financial statements. (For an example of an official statement on this matter see p. 38) Therefore, any model which improves the auditor's ability to establish such a relationship should be more useful to the auditor.

evidence for the accuracy of the model, particularly compared to various other possible approaches. This may be a serious drawback since there are several approaches possible just with the beta-normal model when applied to a "live" system and there is no evidence available about which would be preferred. To add to these implementation issues, it appears Grimlund himself is not certain just how completely an internal control system can be modeled using his theory. He thinks it "unlikely" that in a practical application his approach could be used to completely model an audit environment.[28] However, what is considered far or complete enough is completely left open to question. Thus it appears that although in theory the model is developed, its practicality and accuracy must still be demonstrated. The model does not yet have a field study supporting it as is the case for the reliability model.

There exists a major philosophical difference between the reliability and the Grimlund models. The reliability model puts much more stress on system structure—the relationship of internal control points within a system—than does the Grimlund model. Grimlund has tended to ignore system structure at the subsystem level (i.e., the relationships among the subsystems). For example, in his most recent paper Grimlund argues for using a weighted sum scheme of various error sources in computing system reliability.[29] He apparently makes this

[28] Grimlund, p. 11.

[29] Richard A. Grimlund, "The Integration of Internal Control System and Account Balance Evidence," working paper, University of Iowa, August, 1978, p. 15. Also Grimlund, "A Framework for the Integration of Audit Evidence," p. 88.

simplification for purposes of mathematical expediency; yet one wonders what the ultimate effects of this simplification are on the accuracy of Grimlund's analysis. This is particularly troubling when one considers that the extensive theoretical developments that have taken place in the engineering literature are based to a large extent on the structure of a system. In fact exact system reliability cannot be computed unless the structure function is identified.[30] One would expect Grimlund's approach to be losing some accuracy in letting system structure be determined by a rather artificial weighting scheme. Of course, this issue leads eventually to the relative accuracy criterion of different models which in turn can only be resolved by additional research.

On the other hand, Grimlund's model does accomplish much more than the reliability models by themselves in that it provides a vehicle for the direct assessment of the dollar impact of the system. (Of course, by the addition of an error size assumption the reliability model can also be used to obtain such an error estimate and this prospect is discussed later in this section.) Grimlund even goes so far as to argue that his model is superior to a simulation approach. He bases his argument on cost and accuracy criteria.[31] There exists

---

[30] See Richard E. Barlow and Frank Proschan, Statistical Theory of Reliability and Life Testing Probability Models, (Holt, Rinehart and Winston, 1975), especially chapters one and two which deal with mathematical properties of system structures and their impact on reliability.

[31] Grimlund, "A Framework for the Integration of Auditing Evidence," pp. 49-50.

evidence from other sources that some of these assertions may be correct.[32] Hence, Grimlund's beta-normal model, although not field tested, has as much potential as any other for considerably aiding the auditor in his task.

Reasons for the relevance of reliability
theory for auditor decision making

In discussing reliability theory and the other internal control mdoels, reference has been made to the engineering literature. At this point it appears appropriate to discuss just why an engineering approach, reliability theory, should be relevant for auditors. This may also prove to be useful for a better understanding of the rationale underlying different approaches to modeling internal controls.

The reliability problem as generally stated in the engineering literature is the problem of predicting system performance.[33] Frequently, the prediction is an estimate of the value of a probability, interpreted as "system reliability." More specifically engineers define reliability to mean "the probability of a device (or item or organism) performing its (or his or her) defined purpose adequately for a specified period of time, under the operating conditions

---

[32]Kottas, Lau and Lau provide evidence that a "four moments" approximation to a large class of stochastic management model output variables can result in less computations and more accurate modeling than using Monte Carlo simulation. See John F. Kotas, Amy Hing-Ling Lau, and Hon-Shiang Lau, "A General Approach to Stochastic Management Planning Models: An Overview," Accounting Review, April 1978, pp. 389-401.

[33]The general definition of a system is a set of elements of components which operate together to accomplish an objective.

encountered.[34] By using the very broad definition of system it is readily apparent that there is nothing conceptually wrong with comprehending in the reliability definition the probability associated with any system performance. Therefore, it is apparent that the following definition for use in an auditing context is thus the conceptual equal of the engineering definition: the probability of successful operation of a system of internal controls. Note that this definition leaves a considerable latitude in deciding just what constitutes "successful operation". Some alternative interpretations of this concept are discussed in chapter four.[35]

There are two basic approaches possible in estimating system reliability. One is to test the entire system as a whole. Such a system test in effect amounts to a direct test of the output of the system and there is no additional complexity introduced through the decomposition process. In an auditing context the output of the internal control system is the set of financial records, and thus a substantive test of the records can amount to a direct test of the output of the system (i.e., a direct systems test).

---

[34] Nancy R. Mann, Ray E. Schaefer, and Nozer D. Singpurwalla, "Methods for Statistical Analysis of Reliability and Life Data," (John Wiley & Sons, 1974) p. 1.

[35] Footnote 21 has already given the most common audit interpretation.

The other basic approach to estimating system reliability is to use the data obtained from tests of subsystems or components which make up the system. In engineering applications this is frequently the only approach possible because of expense or time limitations or simply the fact that it is virtually impossible to test the entire system without destroying it. System reliability theory was developed as a response to this problem. In practice the system is decomposed to the component or subsystem levels for which reliability data is available, then the overall system reliability is computed from these subsystem reliabilities via an organizing structure or structure function which is a mathematical representation of the dependence of system reliability on subsystem reliabilities.

Fortunately, auditors are not as constrained as engineers. In fact auditors always have the option of directly testing the output of the system via substantive tests and not worrying about the theory of formally integrating data obtained from subsystem tests. However, the reliability theory does deal with the problem of predicting system performance based on the analysis of subsystem data only. Since auditors usually gain much information about subsystem performance in the process of analyzing organizational records, it stands to reason that for the sake of economic efficiency such information should be used to reduce the amount of system testing (substantive testing) wherever possible. Systems reliability theory indicates the information can be comparable to that obtained from a direct systems test.[36]

---

[36]It seems clear, however, that estimates obtained directly from

In fact this reasoning parallels within the more formal reliability theory framework that which leads to the internal control hypothesis of auditing.

However, in making the transfer of the theory from the engineering framework to the auditing framework, a somewhat subtle dimension is added in the carryover. This is that whereas in engineering the two approaches for evaluating the system reliability are usually mutually exclusive due to the nature of the engineering problem; in auditing, the two approaches are used in combination, one source of information replacing the other via the linkage rules, and only rarely is there complete reliance on only one source.[37] It is this added dimension that introduces the problem unique to auditing, at least in the non-Bayesian sense, of determining the extent to which the information can be interchanged.

To further complicate the picture, it should be reiterated that a system reliability measure is not the measure of system performance of ultimate concern to the auditor. In practice the auditor is always ultimately concerned with fairly measuring financial performance--a variable measured in dollars and cents, not necessarily some

---

system tests will usually be more trustworthy than estimates based on subsystem data, since the latter require the assumption of some mathematical model. Thus, in practice, the two sources of information will not exactly be comparable. There are other reasons peculiar to the audit setting for this noncomparability which will be discussed later in this chapter.

[37] In practice, however, many auditors may never rely on internal controls to reduce substantive testing. This may be particularly true in audits of small business firms. See for example Arens and Loebbecke, p. 179.

probability measure. Thus a complete audit model should incorporate the eventual dollar impact of system performance. This is automatically accomplished in substantive testing by the very nature of the test. However, this is not accomplished by the reliability theory model described so far. Considering that the auditor's ultimate goal is to decide on the fairness of the presentation of the financial statements, rationally he ought to be relating the condition of internal controls to their monetary impact. Thus a reliability theory model as described so far is only a partial model with regard to auditor's decision making.

What is necessary to make the reliability model complete for audit purposes is to introduce an error size assumption to the analysis, and indeed this has been proposed even with the first introduction of the model by Cushing.[38] However, this assumption introduces some rather subtle and more general issues about how to best model an internal control system. These issues can greatly affect the tractability of the mathematics involved. Grimlund appears to be the one who has most thoroughly explored the various avenues to modeling internal controls; although by failing to address the system structure issue more completely, this researcher feels that Grimlund's analysis may have missed an important aspect of the modeling. However, since it is not the purpose of this dissertation to develop new internal models, but rather to help assess the importance of existing ones, no

[38]Cushing, pp. 37-38.

attempt is made to extend Grimlund's extensive work. Instead, a brief survey of the issues is made to help justify the direction the dissertation takes. The issues are considered with the help of a somewhat artificial division of internal control modeling theories into three major approaches: traditional, reliability, and Grimlund. These approaches appear to encompass all of the model types that have evolved so far.

Issues relevant to alternative conceptual approaches to modeling internal controls

The approach that has the most historical support, at least by implication, is one that is referred to here as the traditional approach. This approach recognizes that a relationship exists between the amount of monetary errors in the financial records and certain attributes of the system of internal controls. Typically, these attributes take the form of dichotomous events which can be mathematically represented by a Bernoulli process and so the information consists of binomially distributed pass-fail subsystem data. This modeling is implied by official AICPA pronouncements. The relationships between the attribute data (frequently in the form of compliance error rates) is left undefined except that it is assumed that the auditor through professional expertise can decide when a particular attribute rate leads to excessive monetary errors. Interaction effects among several attributes are typically ignored.[39] The approach amounts

[39] See Roberts for a recognition and discussion of this deficiency of present audit practice, Donald M. Roberts, Statistical Auditing, (AICPA, New York, 1978), p. 147.

to a black box concept where it is left to the auditor's subjective judgment to establish the relationship between internal control attributes and monetary errors.

In light of the many uncertainties that accompany a real audit this approach is understandable. In the real world, especially when behavioral factors are considered, there are probably as many different internal control systems as there are economic entities in operation. A complete mathematical modeling is also constrained by the fact that qualitative as well as quantitative factors and, as a result, subjective judgments play the most important role. Thus to some extent a complete analytical model is not possible. In such an environmental setting perhaps general conceptual guidelines as represented by the traditional approaches may be optimal after all the costs and benefits of the models have been considered. Again the conclusions are dependent on the extent of internal control hypothesis validity.

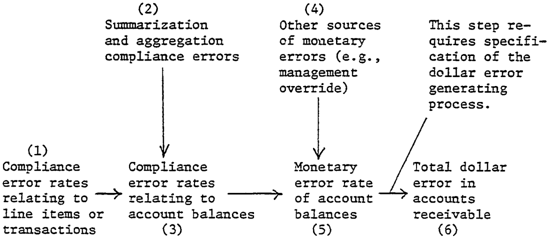
On the other hand, it is likely that a more refined analytical model can improve decision making by limiting subjective assessments to more specific or simpler forms which are easier for the auditor to handle. Thus most of the integration is done formally by the model instead of being left to the auditor's judgment. This implicitly assumes that the auditor's judgment is more informative at a lower level of aggregation.[40]

---

[40] There is a limit to which this statement is true, however. An otpimal level of aggregation is a function of data availability as well as auditor judgment processes. For example, is the auditor better equipped to judge the total error of an account or to judge separately all the component sources of error? A lot, of course,

The three approaches reviewed here essentially differ in terms of the level of abstraction or, as Grimlund calls it, the degree of integration of the associated models. To better highlight the way various approaches can differ in analyzing an internal control system, it is useful to illustrate the relationships with a diagram. Figure 3 illustrates the logical relationships one might expect in an accounting system for accounts receivable. This diagram will be used as the basis of discussion about the three modeling approaches.

Fig. 3. Diagram of Relationships Pertinent to Internal Control Modeling.



---

depends on data availability. Neither of these topics has been sufficiently researched to identify or even indicate an optimal level of aggregation for modeling purposes. Thus it is unclear whether the present traditional approach is optimal or whether, indeed, more refined models can offer significant improvements in auditor decision making.

The traditional audit approach is to expect the auditor to make the leap from (1), (2), and (4) directly to (5) by use of his professional judgment. Considering that (1), (2), and (4) themselves frequently represent several sources of error where qualitative as well as quantitative factors play an important role, this is a tall order for any professional or any analytical model to fill. Perhaps this is why the actual modeling of the integration process has been generally ignored under the traditional approach and the whole subject swept under the rug of professional expertise. However, given the formidable intellectual problems such integration poses, a high potential for avoiding many subjective errors exists if the integration is done objectively using a formal mathematical model. This is the underlying rationale for developing more complete mathematical models.

A more recently evolving approach for modeling internal controls is one that is called here the reliability approach. This approach formally attempts to model the system structure, i.e., the relationship of the underlying controls or subsystems and the data associated with the subsystems, to obtain a reliability measure of the overall system. The reliability can be measured several different ways but the measure that appears to be used most frequently in an audit context is the proportion or probability of records not having a monetary error (note that under this definition reliability also equals 1-[monetary error rate of the system], where monetary error rate is the proportion of all records processed by the system having a monetary error).

A monetary error rate measure for a system (item (5) of figure 3) represents a higher degree of integration and is more potentially useful data to the auditor than just compliance errors (particularly when there are several types of such errors) because the former is more relevant for assessing the numerical accuracy of accounts receivable. An even more useful measure, however, is item (6) because this is the measure of direct interest in deciding whether to accept or reject the accuracy of the reported amounts. Thus the reliability measure, though providing considerable improvement in aiding auditor decision making, falls short of giving the auditor the measure he is directly interested in--the total amount of dollar error.

Now it is readily evident that by making some assumptions about the size of the dollar errors as well as the error rate (1-reliability = error rate = proportion of dollars having monetary errors), the reliability measure can also be very useful in estimating the total amount of error. This possibility is comprehended within the meaning of the term reliability approach as used here. Thus, the reliability approach has the potential for allowing the auditor to more objectively relate compliance error rates with the total amount of dollar errors in the accounts by including an error size assumption.

It should be noted that in spite of considerable research in the area of system reliability measurement by many talented individuals, the problem of measuring system performance for complex systems is "virtually an unsolved problem." This is particularly true when both the system and component subsystem measures of performance are a variables measure such as dollars.[41] For the case where the measure is in terms of a probability associated with the system, a system reliability, methods have been developed for calculating confidence limits using data obtained from tests of subsystems or components. However, for a complex system this is still a largely unsettled area--even when the structure is simple, but there are many components.[42] The literature has produced many alternative models thus indicating that the statistical theory in this area is still evolving--this explains why Stratton had problems estimating certain system reliabilities in his field study.[43] Nevertheless, any model of the internal

---

[41]According to Hillier and Lieberman, "Statistical estimation of component reliability is well in hand, but estimation of system reliability from component data is virtually an unsolved problem." See Frederick S. Hillier and Gerald J. Lieberman, Operations Research (Holden Day Inc., San Francisco, Second Edition, 1974), p. 594. The quote refers to the simpler problem of probability measure, not to the more difficult variables measures. The fact that Grimlund attempts this is what is so significant about his work. He has probably done more theoretical work for implementing what is called here the Grimlund approach than anyone else. Unfortunately, he uses the same methodology when the subsystem data is of the binomial type and clearly here his approach could have benefited more from the reliability literature.

[42]For example, exact confidence bounds for a series systems exist only for a two component system. Close to optimum bounds exist for a three component system; and no nonrandomized (that is, depending only on discrete attribute or pass/fail data) exact bounds have been developed for higher numbers of components in a series system. See Mann, Schafer and Singpurwalla p. 488.

[43]For a discussion of and analysis of the performance of some of the methods available for computing bounds on just the series system reliability see Alan Winterbottom, "Lower Confidence Limits for Series System Reliability from Binomial Subsystem Data," Journal of the American Statistical Association, September 1974, pp. 782-788. Also Mann, Schafer, and Singpurwalla pp. 496-516.

control system that ignores or does not attempt to accurately represent the system structure is probably going to result in serious errors of measurement (see p. 52).

The reliability approach (as distinguished from the reliability model) can be used to formally integrate all audit evidence, attribute and variables data (but only at the systems level for variables); as well as subjective assessments since the approach is amenable for use in a completely Bayesian framework.[44] The reliability approach differs from the traditional approach by the fact that integration of audit evidence is done formally via mathematical models. It differs from the to be described Grimlund approach in the timing of the modeling of the monetary error size generation process. Under the reliability approach this modeling is always assumed to be done after a systems monetary error rate estimate is obtained. That is, the reliability approach assumes the modeling is always a two stage process: first an estimate of system monetary error rate is obtained, then an error size generation process is assumed. The two stages working in conjunction thus yield an estimate of the dollar amount of financial error caused by the system.

---

[44]Reliability theory has also evolved to the point of having developed a Bayeisan framework. See Mann, Schafer, and Singpurwalla especially their chapter eight, "Bayes Methods in Reliability". According to the authors, the largest practical problem associated with implementing the Bayes approach is assignment of costs or utility, with assignment of the prior distribution being a "nontrivial" problem (p. 383). They feel that the main usefulness of Bayes methods is in providing a means of combining previous data with observed data to obtain better estimates of parameters by using the posterior density." (p. 385) However, certain technical difficulties arise in interpreting

The reliability approach includes several of the models Grimlund has developed. In particular it includes all his models in which a monetary error rate estimate for the system is first obtained and then an error size assumption is made.[45] However, as pointed out earlier, for various reasons Grimlund does not consistently use systems reliability theory to obtain an estimate for the monetary error rate generated by the system when the subsystem data is of the binomial form.[46]

The Grimlund approach is thus defined to include all models which attempt to separately measure not only the monetary error rate of each control point or subsystem, but the error size generation process as

---

posterior distribution parameters when variables data (in this case failure times) is used (pp. 396-398).

[45] In fact the reliability approach is implied by the organization of Grimlund's dissertation in that the first two chapters of his theoretical work (chapters four and five of his dissertation) develop methods for integrating error rate information and chapter six then introduces the error size information to the analysis. It is only in chapter five that what has been called here the Grimlund approach is introduced. In addition it should be noted that even the more complete Grimlund approach is limited by the fact that Grimlund assumes all error size processes, regardless of the level of aggregation, can be modeled by the normal distribution.

[46] Grimlund does recognize certain system structures which he uses to model what are called here the subsystems or components of the system. His inconsistency (or perhaps it is a matter of incorrect interpretation on the researcher's part) arises in integrating the subsystem error rate distributions using a weighted average of the beta probability density functions associated with each subsystem (p. 88 of his dissertation). This in turn arises because he uses the method of computing probability moments for the error rates associated with the component accounting subsystems. The end result of this process is that while systems structure is recognized at the lowest level of decomposition it is generally ignored at the intermediate subsystems level. Reliability theory, on the other hand, is more consistent in

well. For example, looking at figure 3 the Grimlund approach attempts to separately model the dollar error generated at (1), (2), and (3) and integrate the process to model the net effect at (5). This approach though conceptually more complete and "realistic" introduces some difficult theoretical problems. One of these is the fact already mentioned that ignoring to an extent the system structure is introducing possibly serious errors. The engineering reliability literature puts great stress on the structure in the methods that have been developed for calculating confidence limits on reliability.[47] In fact is it impossible to predict system performance based on subsystem data unless a mathematical representation is assumed to be given relating relevant system characteristics to the quantities which can be measured in subsystem tests. Grimlund's approach implicitly obtains such a structure representation as a result of an artificial weighting scheme derived from a truncated form of a Jacobi polynomial orthogonal expansion, whereas reliability theory puts more stress on more accurate modeling of the actual system structure.[48]

---

recognizing structural importanct throughout the analysis. Grimlund has given some recognition to these differences (see p. 67 of his dissertation) but it is not clear why he opted for his method of moments approach. Perhaps he felt the method of moments had fewer problems.

[47]This is true for data consisting of variables as well as attribute measures, Bayesian and non-Bayesian methods. See chapter ten of Mann, Schafer and Singpurwalla, especially pp. 465-466.

[48]Grimlund probably does this because he also uses the same Jacobi orthogonal expansion for error size modeling and it is not clear whether with the Grimlund approach some system representation accuracy must be sacrificed when one wants to model variables at the component level.

Other theoretical problems with the Grimlund approach include his error size process assumption, his methods for aggregating error rates and error sizes, and his various alternatives for approximating the sum and difference of mixtures of beta-normal random variables to obtain a summary probability density function (this is necessary in order to consolidate the error rate and error size information from each of the subsystems into an aggregate probability error distribution which can then be used to estimate the amount of dollar error generated by the system).[49] The upshot of these problems is that there is considerable doubt as to the accuracy of the approximations and the theoretical distributions which are approximated. In all fairness to Grimlund, however, it must be remembered that he has very ambitiously charted considerably new territory for auditors in analyzing the mathematical issues associated with integrating all audit evidence within a Bayesian framework. Although he has failed to provide evidence on the relative accuracies of his proposals--and he fully recognizes these limitations--he has developed a potentially improved conceptual approach which may allow auditors to considerably reduce the scope of their subjective judgment and consequently improve the quality of their decision making. Another factor which should be considered in judging Grimlund's work is that relatively little is known about the environments in which auditors operate, consequently, final resolution of some of the issues raised by his research must

---

[49]Grimlund, "A Framework for the Integration of Auditing Evidence," pp. 89-101 and pp. 130-131 concisely discuss all of these issues.

await further empirical work.

One result of the problems associated with the accuracy of Grimlund's formulations is that the more complex the system becomes, the more uncertain is the accuracy of the error estimate. It is perhaps for this reason that Grimlund appears to balk at the notion of attempting to comprehensively model all sources of error in an internal control system and, instead, expects the most usefulness to result from analyzing isolated weaknesses of the internal control system.[50] He also notes the possibility that error size modeling at the aggregate systems level (just as under the reliability approach) may be "particularly reasonable".[51] In fact, there is nothing to prevent separate stage modeling of error rates and error sizes as implied by the reliability approach when using Grimlund's theory. Thus Grimlund's models can fall into the reliability approach category if the appropriate perspective is taken in modeling the accounting process. Thus the naming of these approaches does not necessarily reflect the type of model being used.[52]

---

[50] Grimlund, "A Framework for the Integration of Auditing Evidence," p. 11 and p. 183.

[51] Grimlund, "A Framework for the Integration of Auditing Evidence," p. 93.

[52] Similarly, using reliability theory to model both the error rate and error size for each component (which ultimately may be the most theoretically accurate method) would be considered use of the Grimlund approach as defined here.

The essential difference between the best available model using the reliability approach and the best available model using the Grimlund approach to internal control modeling is twofold: 1) the difference in timing as to when the error size modeling should be attempted (i.e., whether at the subsystem or system level), and 2) the degree of importance attached to an accurate mathematical representation of system relationships. In the first source of difference the Grimlund approach is definitely more "realistic" but, unfortunately, fraught with some analytical problems whose ultimate effect on accuracy is unknown. In the second source of difference, the reliability approach is clearly more "realistic" and has fewer theoretical problems because of the more extensive research supporting it.[53] Thus it appears that, of the three approaches, both the reliability and Grimlund approach are conceptual improvements over the traditional approach. However, it is not clear if either one is more accurate than the other, although there is less potential for theoretical problems under the reliability approach.

Since consistency suggests that a choice be made from among the approaches for the dissertation, other evidence should be considered before deciding on the approach that a priori appears to be best for

---

[53]This assumes that a reliability model using the Grimlund approach, as discussed in footnote 52, is not considered. No one has yet developed such a model, although in concept it may be the best model possible. However, for purposes of this discussion it is not considered an existing model using the Grimlund approach.

the present audit setting. This leads to consideration of the fact that Grimlund's approach may require auditors to significantly change their data collection procedures. For example, present practice puts much stress on collecting binomial type data (compliance error rates) on the internal control system as opposed to variables (error size) type data.[54] Perhaps this merely indicates a deficiency of present audit practice and, therefore, should not be used as a basis for guiding the research. On the other hand, the typical audit may not provide enough data on which one can expect to accurately model the error generating process of each subsystem. This leads to the more general issue of what level of aggregation is the auditor's judgment most informative and for which reliable data is available. For example, in highly reliable systems where the output of the overall system has very few errors, the component subsystems could be expected to have commensuratively fewer errors (assuming the system is not excessively redundant) so that at the component level the errors would be extremely rare. Under such conditions (and this appears to be the norm in audit practice), error size information would be so scanty that modeling of the error size process could be infeasible at the subsystem

---

[54] Part of this emphasis may revolve around the fact that the substantive tests by definition are concerned with measuring dollar accuracy, although there is some recognition that the auditor inspect the "propriety" of internal control performance during the internal control study phase, which implies checking the monetary accuracy of the related record or relating the compliance deviation to the monetary amount of the related record. On the whole, though, the emphasis is on attribute data, see Sec. 32055-.59, in particular paragraph .58, Sec. 320A.22, and Sec. 320B.15. of AICPA, <u>Auditing Standards</u>, also see footnote 3.

level. Grimlund found that even in his case study illustration he had to resort to some simplifying assumptions because of the lack of information at the decomposed level.[55] What effect this has on accuracy can only be conjectured. Thus the degree to which an error size generating process for a subsystem can be modeled is strongly influenced by the complexity of the overall system, the knowledgeability of the auditor, the resources available for the task, and the reliability and extensiveness of the data available on the subsystem. It is these information availability issues as well as cost-benefit factors that might explain the kinds of internal control information auditors have been gathering in practice. If that is the case, then the Grimlund approach as defined here may not be of much use to auditors even if it proves to be more accurate than the reliability approach.[56] Thus, conformity to present practice should play an important role in deciding which approach to use in the research.

Decision on the conceptual approach and internal control model to use in the simulation

After considering these issues the researcher decided that the reliability approach is the more appropriate one to use in the dissertation. The reasons for this are: (1) the reliability approach is

---

[55]Grimlund, "A Framework for the Integration of Auditing Evidence" pp. 119-125. This problem of error size modeling crops up even at the aggregate system level (pp. 130-131 Grimlund) and proved to be a source of difficulty in the simulation study of this dissertation (to be discussed in chapter five).

[56]Grimlund recognizes that the degree of integration of a model is not necessarily a measure of its usefulness, p. 32 of Grimlund, "A Framework for the Integration of Auditing Evidence."

simpler mathematically and avoids certain mathematical problems in application; (2) it is better developed and has more extensive research supporting it because of applications in engineering (the Grimlund approach although more "realistic" in certain respects is untested in terms of the technical problems that may arise because of inaccuracy, mathematical intractability, and data availability); (3) it conforms closer to the modeling implied by present practice and is consistent with the traditional approach; (4) it conforms closer to the kinds of data used in practice (whether due to data availability constraints or otherwise); (5) it is compatible for use with some of Grimlund's models, in particular it is compatible for use with the only published version of Grimlund's work, the Felix-Grimlund model, which is Grimlund's general mixture of beta-normals model simplified for use with a single substantive test; (6) and finally, because of the tradeoffs involved, it may prove to be more accurate than a fully implemented Grimlund approach for modeling internal controls.

The consequences of using the reliability approach in the dissertation are twofold: (1) it considerably simplifies the simulation of the accounting environments and reduces the number of assumptions that need to be made, and (2) it determines how the Felix-Grimlund model is simulated and avoids some mathematical problems associated with using the model. Both of these effects are more completely discussed in the pertinent sections describing the simulation.

The importance of the kind of approach used in modeling an internal control system is considerably reduced by the fact that the upper bound on the value of internal control information is independent of the approach used to obtain that information. The ultimate measure of interest to the auditor is the amount of dollar error in the accounting records--which is the output of the system. Thus for purposes of determining the extent of output testing (i.e., substantive testing) only the potential for the amount of error is of interest and there is no intrinsic value associated with knowledge of the internal controls. Internal controls are thus mainly useful for predicting the output, and the more accurate such a prediction is the more useful it is to determine the amount of direct output testing. It logically follows that the maximum value of internal control information is the complete elimination of the substantive testing that would have been done without such information.[57] This is true regardless of which approach is used to model the internal control system and thus the upper bound is independent of the approach.

This conclusion is further tempered by the following observation pertaining to the present institutional setting. Auditing theory recognizes that there are differences in reliability (in the nonmathematical sense) among the several kinds of audit evidence that are available. In particular, evidence obtained from a source outside of the entity or that which is objective (e.g., direct confirmation of accounts or physical observation of assets, which are forms of substantive tests) is more reliable than evidence obtained from within the entity (evidence on which internal control tests are usually

---

[57]Remembering that the value of information dealt with in this dissertation is that associated with affecting the extent of subsequent audit procedures.

based). This means, generally speaking, substantive tests are more reliable than the tests of the associated internal control system. The auditing profession has evolved a response to this reality by disallowing such extensive reliance on internal controls that substantive testing is eliminated.[58] A certain amount of substantive testing (and in particular tests of details) is thus required but no formal minimum has been established. However, a convention has evolved that at least a 50% statistical confidence level or equivalent should be obtained from the substantive test.[59]

Considering the possibly huge data and/or resource requirements of the various internal control models and the fact that the impact of this information is limited, an unanswered question is how much effort the auditor should spend to learn about the internal controls.[60] That

---

[58]AICPA, Auditing Standards, Sec. 320.71.

[59]The minimum confidence level suggested appears to be influenced by the sampling procedure used. Generally, dollar-unit sampling practitioners do not appear to set confidence below .8 (e.g., the Clarkeson Gordon & Co.) whereas practitioners of classical statistical estimators seem to be more willing to go as low as .5 (e.g., Ernst & Ernst and Peat, Marwick, Mitchell & Co). Even auditing standards imply a low of .5, for their example. See Sec. 320B.35. Certainly the possible dependence of the confidence level on the sampling procedure and evaluation method warrants investigation.

[60]Another of Stratton's survey findings was that 5-30% of total audit effort is expended on compliance testing on a first engagement. See Stratton p. 46.

is, are the models cost-benefit justified? Therefore, it is somewhat surprising to find that of all the internal control model researchers only Grimlund made a serious effort to link the output of his model to the amount of substantive testing. However, even Grimlund concluded that further empirical research is necessary to determine whether use of such models can improve auditor performance compared with less sophisticated techniques.[61] Thus all the models rely on the validity of the internal control hypothesis to justify their use. This then naturally leads to the question of what evidence does there exist for the validity of the hypothesis? The relevant research is reviewed in the next two sections.

2.4 Review of behavioral research relating to internal control evaluation and its impact on subsequent audit procedures

In view of the prominence given to internal control information in auditing theory and in auditing standards, it is surprising how ambiguous the evidence is concerning the validity of the hypothesis in practice. A review of available firm audit manuals indicates that some firms are willing to rely more on internal controls than other firms.[62]

---

[61] Grimlund, "A Framework for the Integration of Auditing Evidence," p. 18 and pp. 129-132.

[62] See footnote 59 for examples. Generally the decision on confidence level with reliance appears fairly subjective. Here is a typical wording: "However, once the auditor decides to conduct a statistical test, he would probably not want to bother with any test which had less than an even chance of discovering a material error. Therefore, the upper limit on β could reasonably be set at .5, giving a range of .05 to .5" (Elliott and Rogers p. 50, see reference in footnote 79).

Anecdotal evidence indicates many auditors are confused about how internal control information should be used and what impact it should have on the audit.[63] On the other hand, some firms are apparently putting a great deal of faith on the validity of the hypothesis to help moderate or cut the steep rise in audit fees.[64] These observations indicate wide inconsistencies on the amount of reliance placed on internal controls. Generally, these indications of auditor inconsis' icies are supported by the findings of more formal studies pertaining to this topic.

According to Weber, early research by Arens consisting of the examination of the working papers of auditing firms to see how various factors affected the audit plan, found that internal control information did not have a significant effect on subsequent audit procedures. Audit procedures tended to be consistent across clients but differed across audit firms.[65] Aly and Duboff found that sample sizes varied considerably (10% to 100% of accounts receivable) when auditors were asked to recommend sample sizes for an audit of a small retail firm.[66]

---

[63] In discussions with practitioners the researcher has gotten the general impression that either, (1) practicing auditors rarely rely on internal controls, or (2) they are not sure of the relationship of internal controls to substantive tests.

[64] See Harvey Kapnick's comments in "Holding the Line of Audit Fees," Business Week, (October 23, 1978), pp. 57-58.

[65] Ronald A.G. Weber, "Auditor Decision Making: A Study of Some Aspects of Accuracy and Consensus" (Ph.D. dissertation, University of Minnesota, 1977), pp. 13-14.

[66] Hamdi Aly and Jack Duboff, "Statistical vs Judgement Sampling: An Empirical Study of Auditing the Accounts Receivable of a Small Retail Store," Accounting Review, January, 1971, pp. 119-128.

Similarly, Corless found that "there was considerable variability among the prior distributions assessed by different auditors for each audit case," in a mailed questionnaire study that investigated the feasibility of applying Bayesian statistics to auditing.[67]

In a controlled experimental setting Ashton investigated auditors' ability to integrate internal control data from an internal control questionnaire designed for payroll, and to judge the strength of the overall internal control system on a six point scale.[68] The results showed that auditors have a high level of agreement on overall internal control evaluation as well as a high level of decision stability (i.e., intrajudge reliability).

In contrast with this result other controlled experiments obtained contrary findings. Burns found that experienced auditors as well as auditing students made significant errors in assessing the amount of dollar errors generated by failures in an inventory internal control system.[69] These results motivated him to develop the simulation approach to analyzing system performance described earlier. (See p. 48.) However, in examining the effects of using this particular auditor's decision aid on practicing auditors, Weber found that there was little relationship between the "size of the possible dollar error in total

---

[67] John Corless, "Assessing Prior Distributions for Applying Bayesian Statistics in Auditing," Accounting Review, January, 1974, pp. 556-566.

[68] Robert Ashton, "An Empirical Study of Internal Control Judgment," Journal of Accounting Research, Spring 1974, pp. 143-157.

[69] Burns, "Computer Simulation: A Tool for Testing the Effectiveness of Internal Control Auditing Methods."

inventories and the extent of substantive testing proposed."[70] This finding indicates that even if the internal control information can be used to accurately estimate the total dollar error in an account, the amount of substantive testing for that account remains largely unaffected. Thus the finding provides some evidence that the internal control hypothesis is not being followed in practice.

Joyce also examined the effects of internal control evaulation on subsequent audit procedures.[71] He found that his group of auditors showed a low level of agreement on the planned number of man-hours allocated to an audit program of accounts receivable for a given amount of internal control information, and a high level of intrajudge decision stability.

Finally, in probably the most comprehensive of the behavioral studies, Mock and Turner also found a low level of agreement among practicing auditors in planning for substantive test sample sizes based on a given set of information about internal controls.[72] However, the strength of the internal controls does have a marked effect on the nature and variability of the auditor's responses.

In sifting through these findings it is important to recognize that there are several factors which may account for the varying results. First, the kind of judgmental tasks involved in the studies

---

[70] Weber, p. 167.

[71] Edward J. Joyce, "Expert Judgement in Audit Program Planning," Empirical Research in Accounting: Selected Studies, 1976, pp. 29-60.

[72] Mock and Turner.

using subjects varied from study to study. For example, Ashton, Burns, and Corless asked their subjects to make judgments relating to evaluation of internal control, whereas Joyce, Aly and Duboff, Weber, and Mock and Turner asked that the subjects make audit plan decisions based on an evaluation of internal control. Only Ashton's study found a high level of agreement (consensus) among auditors, and this perhaps resulted from the fact that the subject auditor's task consisted of working with internal control questionnaires similar to those extensively used in practice. On the other hand, Burns and Weber found that auditors have problems judgmentally integrating audit evidence to obtain a total dollar error estimate for the internal control system--something which theoretically they should have extensive experience in doing (but which may not be the case in practice as discussed later in this section). A similar disparity from the Ashton finding observed by Corless, however, may have been due to the Bayesian methodology used, which is not common in audit practice. Thus the results of behavioral studies dealing with the evaluation of internal control indicate that for certain tasks (e.g., integration and interpretation of compliance test data) there may exist large variability in auditor judgment (due to such mediating variables as auditing firm effects, auditor experience, risk levels, environmental factors besides internal controls, and other factors discussed more thoroughly in the Mock and Turner Study but will not be pursued here), which in turn can influence subsequent audit procedures and thus impact on the validity of the internal control hypothesis.

The Joyce, Aly and Duboff, Mock and Turner, and Weber studies all have findings pertaining to the behavioral impact of internal control information on audit plan decisions. These studies thus provide more direct evidence on whether auditors behave in conformity with the internal control hypothesis. In general the evidence is ambiguous. The Aly and Duboff, and Joyce Studies found a high variance among the auditors' judgments implying little impact of internal control information. Weber found that when the quality of information improved, due to the introduction of the simulation decision aid as an experimental treatment, there was no effect on planned subsequent audit procedures thus also implying little impact of the internal control information. On the other hand, the Mock and Turner Study which utilized a much more complex design than any of the above (in particular they were the only ones to vary the strength of internal controls as an experimental treatment) found that the subject practicing auditors did systematically recommend smaller sample sizes under the "strong" internal control treatment. A rather disquieting aspect of their results, however, is that the practicing auditors planned sample sizes dramatically different from the "normative" ones the researchers expected. Typically, the average sample sizes were either one half or twice the size of the normative ones. This incongruity was not explained by the researchers. Thus the validity of the implied linkage rules used may be called into question. Nevertheless, the Mock and Turner research is notable for providing the only direct evidence of the behavioral validity of the internal control hypothesis.

In summary the behavioral studies have not provided conclusive evidence on the validity of the internal control hypothesis of auditing. It appears that in practice auditors are not as convinced of the validity of the hypothesis as auditing theory and standards imply. What could be the reasons for this? Perhaps, as the studies indicate, there is too much judgmental error both in internal control evaluation and in subsequent audit planning to allow for a valid reliance in practice. Weber has raised interesting questions about the use of SAS 1 Section 320 as a normative expression of linking internal control evaluation to audit testing. He suggests that Sec. 320 may not be generally applicable, that it may not be operational, and that it may not be cost effective--in short "it may not provide cost effective guidelines for determining the audit plan."[73] He also found evidence that auditors typically do not formally articulate the possible dollar error that an internal ctonrol system may generate and that significant differences exist among auditing firms in the emphasis put on the use of quantitative tools as opposed to qualitative judgements.[74]

Given these uncertainties about the behavioral validity of the internal control hypothesis (i.e., do auditors behave as if the hypothesis is valid), the question that logically arises is whether this uncertainty can be explained by other than judgmental errors. This possibility, then, leads to a consideration of the available evidence

---

[73]Weber, pp. 3 and 4.

[74]Weber, pp. 178-184.

on the statistical validity of the hypothesis. This evidence is reviewed in the next section.

2.5 Review of statistical sampling research in auditing

The statistical validity of the internal control hypothesis of auditing is dependent on the statistical validity of the linkage rule as well as of the estimate obtained from the substantive test. Although there is a fair amount of audit research on the latter, there is hardly any to speak on the former. The only empirical work done on linkage rules that the researcher is aware of was by Kinney.[75] He used three multiplicative forms in his analysis which was based on a largely arbitrary set of conditions. Kinney found that the optimum act in a decision theoretic framework was very sensitive to the functional form of the linkage rule and so recommended more research related to this topic. This sensitivity provides empirical evidence that the validity of the internal control hypothesis may be greatly dependent on the linkage rule used. Unfortunately, his linkage rules did not appear to correspond to anything that is being used in practice. This and the fact that his parameter values were largely arbitrary and that he ignored problems that might arise with the statistical estimator, clouds the significance of his conclusions; although, as a general conceptual model, his proposals may be useful for helping auditors formalize their analysis.

---

[75] William R. Kinney Jr., "Decision Theory Aspects of Internal Control System Design/Compliance and Substantive Tests," Journal of Accounting Research Supplement 1975: Studies on Statistical Methodology in Auditing, pp. 14-29.

Evidence on the performance of statistical estimators under audit conditions is more extensive and this will now be reviewed in a largely chronological order.

The statistical research in auditing of the 1950's and 1960's primarily consisted of introducing classical statistical techniques to auditing. Apparently, the driving force behind the increasing interest in statistical sampling was and is the greater objectivity obtained by statistical decision making, although the efficiency effect may have played a role as well.[76] During this period, there did not appear to be widespread recognition of the special problems which may be encountered in sampling accounting populations under auditing conditions. Much of the literature was based on techniques developed for survey sampling. However, there were some attempts during this period to recognize the special auditing nature of the problem. Neter[77] appears to be the first to have suggested a link between precision limits and/or confidence level and the evaluation of internal controls. He also showed that sample size can be very sensitive to a change in precision limits or confidence level. His work may have spawned subsequent efforts during this period to develop models which, it was

---

[76] See Grimlund, "A Framework for the Integration of Auditing Evidence," pp. 20-22. Also, in a discussion with one partner of a regional firm, the researcher learned that one common form of judgmental sampling (at least on bank audits) is to always examine a specific proportion of the population (e.g., 5%). This can result in a very inefficient form of testing for large populations. See Herbert Arkin, Handbook of Sampling for Auditing and Accounting, (McGraw Hill Book Co., Second Edition, 1974) pp. 10-11.

[77] John Neter, "Applicability of Statistical Sampling Techniques to the Confirmation of Accounts Receivable," Accounting Review, January 1956, pp. 82-94.

hoped, would help put the audit process on a more objective basis. Examples of these efforts include the attempts by Matuz, Mini, Carmichael, Swieringa, and Brown to develop more objective internal control models; and the introduction of Bayesian revision procedures to compliance tests by Kraft, Tracy, Sorensen, Knoblett and Corless.[78] However, these were all piecemeal appraoches because little was done to integrate these approaches with other stages of the audit process. There was implied acceptance of some form of the internal control hypothesis, but there was no attempt to articulate the hypothesis further or measure the impact on substantive testing.

Research activity increased dramatically during the 1970's when it became more widely recognized that there were special problems associated with statistical sampling of accounting populations. This has been given some official recognition in SAS No. 1, as reflected in Sec. 320A.15. This same SAS No. 1, Sec 320 introduced the conceptual relationship between internal control reliance and substantive testing discussed earlier. Thus, for the first time, knowledge of the internal control system was linked to substantive testing by means of

---

[78] R.K. Mautz and Donald Mini, "Internal Control Evaluation and Audit Program Modification," Accounting Review, April 1966, pp. 283-291; Robert Swieringa and D.R. Carmichael, "A Positional Analysis of Internal Control," Journal of Accountancy, Feb. 1971, pp. 34-43; D.R. Carmicahel, "Behavioral Hypothesis of Internal Control," Accounting Review, April 1970, pp. 235-245. R. Gene Brown, "Objective Internal Control Evaluation," Journal of Accountancy, Aug. 1968, pp. 49-56, John A. Tracy, "Bayesian Statistical Methods in Auditing," Accounting Review, Jan. 1969, pp. 90-98; James E. Sorensen, "Bayesian Analysis in Auditing," Accounting Review, July 1969, pp. 555-561; J.A. Knoblett, "The Applicability of Bayesian Statistics in Auditing, Decision Sciences, July-Oct. 1970, pp. 423-440; John C. Corless, "Assessing Prior Disbributions for Applying Bayesian Statistics in

an explicit formula. At about the same time, another linkage was proposed by Elliott and Rogers which explicitly incorporated both α and β risks in the analysis.[79] According to Elliott and Rogers, control of both risks is necessary to meet the overall audit objection of making reasonably certain that financial statements examined by the auditor are materially correct.[80] (SAS No. 1, Sec 320 only considers the β risk).[81]

Auditing," Accounting Review, July 1972, pp. 556-566.

[79]Robert K. Elliott and John R. Rogers, "Relating Statistical Sampling to Audit Objectives" Journal of Accountancy, July 1972, pp. 46-55.

[80]It is important to note how these risks were defined by Elliott and Rogers because these definitions appear to have been accepted by the auditing profession with perhaps an insufficient consideration of their implications. α risk = the risk of rejecting perfectly correct financial statements (i.e., zero errors); and β risk = the risk of accepting financial statements in error by exactly a material amount (p. 47 of Elliott and Rogers). Some reflection on the definition of α risk may give rise to concern as to actually what risk is being controlled for. Using the above definition, only the α risk associated with a population having zero errors is being controlled; i.e., when no errors are found in the sample. Now this may seem strange to the uninitiated, but due to what some statisticians call "sample bounce" (i.e., variability in the sample standard deviation) it is perfectly possible to pick a variables sample, find no errors and reject the population book value. Needless to say, under such conditions both the auditor and client would feel very concerned about the usefulness of the sampling procedure. Yet this is the only type of error controlled for by the above definition of α risk. Some sampling methods, notably dollar-unit sampling, have zero α risk under such conditions. The fact that this is the definition of α risk used for variables sample planning program purposes in auditing has important implications for the control of Type I error when there are monetary errors in the population of less than a material amount. These implications are discussed in chapters four and five and in appendix VII.

[81]See p. 35.

It was during the first half of the 1970's that the more refined internal control models by Yu and Cushing were developed. In addition, efforts were put into developing normative models for auditor decision making. Two of these models were Ijiri and Kaplan's model for integrating sampling objectives in auditing, and Kaplan's stochastic model for auditing.[82] Both of these models concentrated on variables sampling methods and, furthermore, they tended to restrict themselves to only one class of variables techniques, auxiliary information estimators.[83] Neither model explicitly linked the auditor's evaluation of internal control to the substantive testing (i.e., account balance testing). Hence neither model is complete with respect to the audit process. However, a major contribution by Kaplan was to identify the statistical problems associated with estimating accounting populations.[84] Basically these problems revolve around the validity of using

---

[82]Yuji Ijiri and Robert S. Kaplan, "A Model for Integrating Sampling Objectives in Auditing," Journal of Accounting Research, Spring 1971, pp. 73-78; and Robert S. Kaplan, "A Stochastic Model for Auditing," Journal of Accounting Research, Spring 1973, pp. 38-46.

[83]Closely related to the auditing notions of compliance and substantive testing are the statistical concepts of attributes and variables sampling, respectively. Attributes sampling deals with estimating the frequency of items in a population having a certain attribute. Each item either has the attribute of it does not. Variables sampling, on the other hand, deals with estimating the average or total of some variation in measurement (e.g., amount in dollars) possessed by every member of a population. Unlike attributes, there may exist an infinite number of possible measurements which may arise for a sample item.

[84]Robert S. Kaplan, "Statistical Sampling in Auditing with Auxiliary Information Estimators," Journal of Accounting Research, Autumn 1973, pp. 239-258.

the noraml distribution in constructing confidence intervals with typical audit samples sizes when they contain few errors and/or come from highly skewed populations.

Beck analyzed the performances of the difference, ratio, and various regression estimators in an attempt to identify an optimal auxiliary information estimator for audit use.[85] Although he found that his proposed two regime regression model had lower reliability than ordinary regression and thus did not represent an improvement in the performance of auxiliary information estimators, his study did contribute to statistical research in other ways. First, although he found that the difference estimator generally outperformed other auxiliary information estimators, he also concluded that the stratified mean-per-unit and possibly dollar-unit sampling estimators are probably even better performers.[86] These conclusions overall are not inconsistent with those reached in the Neter-Loebbecke Study which will be described later. The most important contributions of Beck's work, however, lie in his more thorough (than the Neter-Loebbecke Study) analysis of the effects of error patterns (understatement, offsetting,

---

[85] John Paul Beck, "A Critical Analysis of the Regression Estimator in Audit Sampling," (Ph.D. dissertation, University of Texas, 1977). Auxiliary information estimators are those statistical estimators which use information on another variable correlated to the variable being measured (audit value). In accounting the auxiliary information comes in the form of the actually recorded values (book value).

[86] Beck, p. 196 and p. 201, where performance is measured in terms of reliability (the proportion of correct confidence intervals compared to the planned confidence level), sampling distribution statistics, and correlation of estimated population mean and standard deviation.

overstatement), error rates, and book value distributions on the performance of classical statistical estimators. His major findings were (1) that the error rate was probably the most important environmental factor impacting on the performance of the auxiliary estimators, (2) that the distribution of the book value population is important in affecting the reliability of an estimator only when the monetary errors are non-offsetting and non-random (i.e., proportional to the book values), (3) that high correlations between the estimated population value and the standard error did not necessarily imply low reliability of the estimator (the significance of this finding is mentioned in chapter one), and (4) that the logarithmic transformation of the data did not appear suitable for use in auditing.[87]

Scott was the first to attempt to apply Bayesian statistical decision theory for linking the auditor's prior assessment of internal control and his sample design.[88] In many respects Scott's objective is very similar to Grimlund's, however, the difference in underlying assumptions considerably weakens the case for Scott's approach. For example, Scott failed to specify the link between internal control information (e.g., compliance error rates) and substantive testing. Thus his model is not nearly as complete as Grimlund's in integrating all audit evidence, and therefore lacks the completeness to yield

---

[87] Beck, see p. 169 for item (1); p. 197, p. 182, and p. 165 for item (2); p. 194 for item (3); and p. 132 and 198 for item (4).

[88] William R. Scott, "A Bayesian Approach to Asset Valuation and Audit Size," Journal of Accounting Research, Spring 1973, pp. 304-330.

insights about optimal audit decision making. It is also difficult to accept the validity of some of Scott's assumptions which appear to be guided more by mathematical tractability than by accuracy of representation.[89] Grimlund has pointed out the practical problem of obtaining sufficient data to validate the multivariate central limit assumption that is involved in Scott's multivariate sample unit consisting of the total daily net error for a set of accounts.[90] Finally, Scott put considerable stress on obtaining the general form of the auditor's loss function, but of course such an approach always raises issues which probably can never be completely settled. In spite of these shortcomings Scott's model is a pioneering attempt to develop a framework for considering many of the diverse issues of auditor decision making with sampling.

Kinney, in another paper, used decision theory analysis to extend the classical hypothesis testing approach in auditing.[91] He also stressed the issue of auditors' loss functions but it is interesting to

---

[89] Scott, see his assumptions on p. 311 and pp. 316-317. The most unrealistic one is assuming that book values are normally distributed. All the available evidence indicates that, on the contrary, accounting book value populations are very highly skewed, which, in part, accounts for the problems associated with statistical estimation.

[90] Grimlund, p. 38.

[91] William R. Kinney, Jr., "A Decision-Theory Approach to the Sampling Problem in Auditing," Journal of Accounting Research, Spring 1975, pp. 117-132.

to note that he used a different loss function from that of Scott.[92] After making a loss function assumption and rather arbitrarily specifying key parameter values (e.g., costs of taking a sample and costs associated with making Type I and Type II errors), he simulated planned sample sizes comparing certain heuristic approaches with his decision theoretic approach, which, of course, had the lowest expected cost. He ignored the issue of error in specifying the amount of reliance on internal control, implying the auditor could exactly specify the probability of immaterial error based on internal control information. In spite of this, his results indicate an auditor would never reduce sampling confidence below .95, thus implying the auditor would rarely reduce substantive testing.[93] This brings into question the validity of his parameter values and/or his form of the auditor's loss function. This problem highlights the difficulties associated with a simulation which uses a specified loss function. While in theory it is necessary to identify a loss function for a complete analysis, in practice so many additional assumptions need to be made about the parameter values and functional form that the simulation ends up losing much of its generality. That is, the results of the simulation become loss function dependent. Since only a finite number of such loss functions can be tested via a simulation, any conclusions pertain only to

---

[92] Scott concluded that an auditor's loss function is probably asymetric (p. 325), whereas Kinney uses a symmetric loss function (p. 131). Why the difference is not made clear.

[93] In particular, see table 3 and table 1 of Kinney, "A Decision-Theory Approach to the Sampling Problem of Auditing."

the limited set of loss functions and parameter values actually tested. Since the present state of the art is not sufficiently developed to reach general agreement on loss functions (e.g., see discussion on p. 21 of chapter one), it appears that the value of simulations incorporating explicit loss functions will be limited.

Yet another problem of Kinney's approach is that he ignored the effect that a sampling procedure used for substantive testing (e.g., stratified mean-per-unit, ratio, or dollar-unit sampling) has on the reliability (in the sense of the probability the process will perform as intended) and efficiency of the audit. This can be a serious problem in assessing how much substantive testing to do since it now appears there exist considerable differences in the effectiveness of different sampling techniques in different environments.

Cushing developed a continuous state representation Bayesian model which is similar to both the Kinney and Scott models.[94] Using a computer simulation, he conducted a sensitivity analysis of the effect of misspecification of parameter values on total audit costs. The most surprising of his findings was that expected total audit costs do not appear to be very sensitive to the optimal sample size used in substantive testing. This result appears to hold regardless of the amount of prior information the auditor has about the population being

[94]Barry Cushing, "Decision-Theoritic Estimation Methods in Accounting and Auditing: Models and Tests," unpublished paper presented at the University of Wisconsin, Dec. 2, 1977.

sampled.[95] However, certain limitations of his approach should be kept in mind in assessing his results. For one thing, it should be noted that his results are also loss function dependent. Thus, although Cushing considered two very frequently proposed functional forms in economic and accounting literature, a linear loss and a quadratic loss function, and he used widely varying parameter values for these functions, the simulation mode of analysis allows the consideration of only a finite number of cases, and the results are limited to these situations. The use of other functions or parameter values may lead to different results. Another limitation of Cushing's analysis is that he considered changes in parameters one at a time only. His simulation did not assess interaction effects on total audit costs from multiple parameter changes. Since his model required the specification of at least nine parameter values, this may have been a serious omission. Another weakness of Cushing's approach was that the analysis restricted itself to the problem of sampling for substantive testing only. No attempt was made to link internal control evaluation to the auditor's prior assessment of the distribution of the dollar values of an account balance. Instead, he just assumed the auditor somehow obtained such a prior assessment. He also ignored the effect on total audit cost of the kind of sampling method used. Thus his analysis suffers from being incomplete with respect to several basic matters which impact on the audit process. Finally, it should be noted

[95]Ibid, pp. 47-48.

that Cushing's simulation did not consider the impact of making a normal distribution assumption for the sample mean estimators. The normal approximation may be a poor one for typical audit sample sizes from highly skewed populations. This introduction of error in the variables estimation process should have been considered in measuring total audit costs. For that matter, Cushing considered only one accounting population of unspecified distribution and error rate. Hence his results may be dependent on using a nonrepresentative population.

Felix and Grimlund have published the most comprehensive Bayesian model yet for assisting the auditor in integrating his compliance and substantive testing.[96] Continuous ranges of values for error amounts and error rates are assumed throughout and the analysis can be made complete with respect to the basic audit process for a single account. A problem is that certain distributional assumptions about accounting populations are made to make the model mathematically tractable. In addition approximations are resorted to at certain stages in the analysis. A major question, then, relating to using this model is how good an approximation are the resulting modeled distributions to actual distributions faced by the auditor? More basically, and this question pertains to all models, is the model cost-benefit justified? That is, are the additional costs of using this more complex model more than

[96] William L. Felix Jr., and Richard Grimlund, "A Sampling Model for Audit Tests of Composite Accounts," Journal of Accounting Research, Spring 1977, pp. 23-40.

compensated by greater efficiencies and accuracies introduced into the audit program? The question reduces to one of relative efficiencies of different models. This issue has not yet been addressed empirically.[97]

It has been noted earlier that auditing populations and environments have their own unique characteristics. This has made it necessary to experiment with different types of sampling methods. Some of the statistical estimation methods that have been proposed include the mean-per-unit and stratified mean-per-unit estimators; the auxiliary information estimators: difference, regression, ratio, and the associated stratified estimators; various combinations of the preceding (e.g., combined mean-per-unit and difference estimator); and, what will be called in this paper, the dollar-unit sampling (DUS) estimator. The DUS procedure represents a more recent response to the special problems of sampling in the auditing environment. It differs radically from the other methods (which are frequently referred to as classical statistical estimators) in the way it defines the population of items to be sampled. Under DUS, individual dollars are sampled from the

---

[97] In fact, none of the decision theoretic models have been tested under "live" or fairly realistic conditions, i.e., with actual sampling taking place for both compliance and substantive testing on a representative accounting environment. Thus there is no indication of the relative performance of these methods, particularly in comparison to the more traditional techniques. Hence the usefulness of the newer methods and, therefore, their value is unknown.

population of interest (i.e., proportional sampling is performed).[98] This is in contrast to the classical estimation methods which sample entire accounts (random sampling) from the population.[99] This difference has the potential for introducing efficiencies and accuracies in the estimation procedure, particularly for the zero and low error rate populations which auditors frequently encounter.[100]

Teitlebaum can be credited for providing most of the theoretical and empirical basis underlying the DUS method. By taking the point of view that the "basic objective of an audit sampling plan is to determine the maximum extent of potential error that may be present in the

---

[98]The basic difference in dollar-unit sampling is that instead of each account or record having an equal chance of selection, each dollar in the population has an equal chance of selection. Of course, each individual dollar selected is not verified by itself. Rather it acts as a hook and drags a whole account balance with it. The account is then confirmed and any errors are interpreted in terms of the impact on the selected dollar. For example, if it is found that an account is 10% overstated, it is assumed that a dollar selected from that account is 10% overstated. Statistical conclusions are then reached on the sample of dollars so evaluated.

One can think of dollar-unit sampling as a limiting situation where stratification by book value yields no further gains since all sample units have the same book value (i.e., one dollar). This effectively results in sampling proportional to the book value of the records.

[99]In classical estimation, the total amount of error or total audit value of each selected record is usually the basis for computing the sample statistics.

[100]The maximum stratification feature discussed in footnote 98 has the potential for using the smallest sample size for controlling a given risk level. However, the theory for integrating variables with attribute data is incomplete and hence some question remains as to what the actual sampling risks are. See James L. Goodfellow, James K. Loebbecke, and John Neter, "Some Perspectives on CAV Sampling Plans," Part I, CA Magazine, Oct. 1974, pp. 23-30; Part II, CA Magazine, Nov. 1974, pp. 46-53.

audited population, at a given confidence level," he has developed estimation techniques for DUS which control this risk at the specified level assuming the worst possible error pattern.[101] This concentration on controlling the β risk raises the possibility of increased false alarm risks (α risk) and much of Teitlebaum's research has gone into developing evaluation techniques for reducing the α risk. The two main techniques he has developed which appear to be used in practice are the tainted attribute random sampling evaluation (TARS), sometimes also referred to as the Stringer bound when only errors of overstatement occur, and the less conservative (i.e., less likelihood of committing a Type I error) tainted attribute cell selection (TACS) evaluation.[102]

Both of these evaluation methods assure the control of the Type II error for a discovery sample size or larger (as Teitlebaum has proven analytically), however, the associated α risks of these procedure are much more difficult to predict in a situation where the error sizes and rates are allowed to vary.[103] This is particularly true when understatement errors can also occur because there is no general statistical theory supporting the TACS and TARS treatment of

---

[101]A.D. Teitlebaum and C.F. Robinson, "The Real Risks in Audit Planning," Journal of Accounting Research Supplement 1975: Studies in Statistical Methodology in Auditing, pp. 73-74. It should be noted that this assumption is not inconsistent with SAS No. 1, Sec. 320.

[102]A.D. Teitlebaum, "Dollar-Unit Sampling in Auditing." Paper presented to the National Meeting of the American Statistical Association, December 1973, pp. 14-25.

[103]Ibid., appendix I and appendix II.

understatements (except for some simulation results obtained by Teitlebaum). Theoretical analysis is difficult because the procedures involve combinations of attributes and variables principles. Nevertheless, Teitlebaum and various practitioners of the DUS methods appear to have found it to be a very useful and effective method for application in practice. Some DUS advocates even go so far as to argue that the method is superior to any classical estimator regardless of the nature of the population being sampled.[104] Whatever the merits of these arguments, it is clear that the DUS approach represents a formidable alternative to the use of the more traditional statistical estimators.

Besides the TACS and TARS methods for evaluating dollar-unit samples, another method widely used in practice is the cumulative-monetary-amounts sampling (CMA) method developed by the auditing firm of Haskins & Sells. The chief difference between TACS or TARS, and CMA is the treatment of understatement errors. Teitlebaum provides evidence that CMA is more conservative than either of his methods, hence, it will not be considered further in this paper.[105]

Some recent research in DUS methods have attempted to assess the relative performance of various DUS alternatives or to develop new DUS methods for better controlling the α risk. (It appears that everyone

---

[104]For example, see R.J. Anderson and D.A. Leslie, "Discussion of Considerations in Choosing Statistical Sampling Procedures in Auditing," Journal of Accounting Research Supplement 1975: Studies in Statistical Methodology in Auditing, p. 59.

[105]Teitlebaum, Appendix III pp. 22-26. It should be noted that other empirical research has concentrated on the Teitlebaum bounds.

recognizes the validity of the Teitlebaum proofs concerning the TACS and TARS methods to always control the β risk within the planned level, and so the major unresolved issue concerning DUS is which method(s) best hold down the α risk.) Neter, Leitch, and Fienberg have developed a tighter bound (implying less conservatism because there is less chance that a Type I error will be made) for DUS based on the multinomial distribution.[106] However, their method is capable of handling a maximum of seven errors (this is frequently exceeded in practice according to Teitlebaum, McCray, and Leslie) and no mention is made on how to obtain a net bound on total error.[107] Thus the method does not appear ready for use in practice.

Garstka developed a compound Poisson model for also computing a less conservative error bound using DUS.[108] However, since Garstka doesn't claim to have computed an exact bound as is the case with the Neter, Leitch, and Fineberg bound, and since, again, no attempt was

---

[106] Stephen E. Fienberg, John Neter, and R.A. Leitch, "Estimating the Total Overstatement Error in Accounting Populations," Journal of th the American Statistical Association, June 1977, pp. 295-302; or John Neter, R.A. Leitch, and Stephen E. Fineberg, "Dollar Unit Sampling: Multinomial Bounds for Total Overstatement and Understatement Errors," Accounting Review, Jan. 1978, pp. 77-93.

[107] A.D. Teitlebaum, J.H. McCray, and D.A. Leslie, "Approaches to Evaluating Dollar-Unit Samples" Paper presented to the AAA Convention, Aug. 1978, p. 12.

[108] S.J. Garstka, "Computing Upper Error Limits in Dollar-Unit Sampling," Journal of Accounting Research, Autumn 1977.

made to develop a method for computing a net error bound (which is automatically done by the classical statistical estimators), it appears that this model is not yet suitable for use in practice either. On top of this is the uncertainty associated with the amount of improvement, if any, provided by Garstka's model to, say, a TACS bound which has proven itself in practice.

Teitlebaum, McCray, and Leslie have attempted to improve on the TACS bound (again by reducing it so as to eliminate some of its conservatism) by developing their own multinomial and Poisson bounds.[109] They provide evidence that their multinomial bound is in perfect agreement with the Neter, Leitch, and Feinberg bound when all errors in the sample have the same tainting percentage. They go on to develop a close Poisson approximation to their multinomial bound which is preferable for practical use because it is independent of sample size and avoids the need for voluminous tables. Finally, they outline a more general Poisson Cell model which can be developed for any number of differing sample errors. However, it is interesting to note that in illustrating the use of this general Poisson-Cell model the bound proves to be higher than one using the original TACS rule. By their remarks it is evident that their new model is not yet ready for general use.[110]

Reneau performed a simulation analysis of the performance of

---

[109] Same as reference in footnote 107.

[110] Ibid., p. 30.

various DUS evaluation methods.[111] He used populations having understatement as well as overstatement errors to compute two sided bounds (one for overstatements and one for understatements); however, he did not compute a net bound on total error with the various methods. His major result was that TACS and TARS was "more desirable" than the other methods.[112] However, he did not find TACS to be more reliable than TARS, something which Teitlebaum analytically proved. In fact, Teitlebaum, McCray, and Leslie have pointed out certain inconsistencies in Reneau's results concerning the TACS method which clouds the validity of that part of his simulation.[113]

Venecek developed and empirically analyzed a Bayesian model for DUS.[114] Unfortunately, his model is geared to the simpler but more conservative TARS evaluation procedure. A computer simulation on the performance of his Bayesian DUS technique for a variety of populations having understatement as well as overstatement errors, and for net as well as gross error bounds found that: (1) TARS and the Bayesian DUS performed fairly reliably under all environmental conditions and in particular "no portentous effects were detected relative to the

[111] J. Hal Reneau, "CAV Bounds in Dollar-Unit Sampling: Some Simulation Results," Accounting Review, July 1978, pp. 669-680.

[112] Ibid., p. 679

[113] Teitlebaum, McCray, and Leslie, pp. 35-36 (their footnote 13).

[114] Michael T. Vanecek, "Bayesian Dollar-Unit Sampling in Auditing," (Ph.D. dissertation, University of Texas, 1978).

existence and treatment of understatements"; (2) the Bayesian DUS performed only marginally better than TARS and this was in terms of reducing conservatism (α risk) somewhat; and (3) the DUS method as a sampling procedure can bias the characteristics of the sampling distribution.[115] The last result appears to arise primarily because Vanecek failed to treat the top stratum items separately as is done in practice.[116] Considering Vanecek three sample sizes of 100, 200, and 300 items, it is evident that most of the time top stratum items did exist in the population.

The results of this research pertaining to DUS indicate that at the present time the best performing general DUS method is still the TACS method. The TACS method can be used in all kinds of accounting conditions and with any number of sample errors. The TACS method asures control of the β risk at the planned level yet at the same time has the smallest, or close to it, actual α risk of any DUS evaluation method. It has been tested in practice as well as in empirical research and is probably better performing than the only Bayesian DUS model that has been proposed to date. Thus, of the many DUS evaluation methods that have been proposed, it appears that TACS is the most defensible of the DUS alternatives available, and so this is the DUS evaluation method used in the dissertation.

---

[115] Vanecek, p. 270.

[116] Top stratum items are those which are always selected due to the systematic sampling nature of the DUS procedure. The items which have a 100% chance of selection should be evaluated separately and combined with the statistical results of the remaining items. If this is not done (i.e., they are treated as if they are not always being

A serious problem associated with all the empirical work that has been done on the performance of statistical estimators so far is the use of the reliability and precision of the estimator as the chief measure of performance. Here reliability is the proportion of times the statistical estimator produces confidence intervals which contain the true audited population value, and precision relates to the width of the confidence interval. As discussed in chapter one the primary disadvantage of using reliability and other indirect measures of the propensity of a statistical estimator to commit Type I and Type II errors, is that these measures do not specify the frequency of such errors for different amounts of accuracy of the financial records.[117] Considering that the indirect measures can result in erroneous conclusions being reached as documented in chapter one, it is important to obtain more direct measures of actual α and β risks; particularly since these are the ultimate measures of interest to auditors. Thus this study will directly measure the actual α and β risks of statistical estimators as well as the usual other statistics.

In an effort to sort among the many different statistical techniques available (DUS as well as classical statistical estimators), the AICPA sponsored a research program headed by John Neter and James

---

selected) the resultant statistical conclusions can be biased by the fact uncertainty is assumed for errors that are certain of being discovered.

[117]The fact that statisticians in the literature reviewed cannot agree on the implications of these measures for comparisons of Type I and Type II risks, indicates that more direct measures would be useful.

Loebbecke to test the properties of a large number of statistical techniques on a variety of accounting populations. The report on this effort was published as AICPA Auditing Research Monograph No. 2.[118] One of the important results of the Neter-Loebbecke Study was that they found their version of the DUS method did not outperform the other methods over all populations and error types.[119] In fact, no single method was found to dominate in all situations; although stratified mean-per-unit did dominate in almost all of them and performed fairly well in the rest.

Based on these results Neter and Loebbecke then developed a decision flowchart to aid the auditor in choosing a sampling method most suitable to the particular population at hand.[120] This approach reflects a basic change in the philosophy that there may exist a single "best" sampling method for use in all audit environments. However, their scheme requires the auditor to be familiar with as many as six different sampling methods. This Neter and Loebbecke decision flowchart approach, as well as the supporting study, was rigorously criticized by Anderson and Leslie, two DUS advocates, on the grounds

---

[118] John Neter and James K. Loebbecke, "Behavior of Major Statistical Estimators in Sampling Accounting Populations," (AICPA, New York, 1975).

[119] Outperformed in the sense stated . . nominal confidence level is less different from actual confidence (as measured by the simulation), and/or smaller relative standard error (i.e., a tighter precision) is obtained. Please see pp. 24-32 of the Neter-Loebbecke Study (reference in footnote 129) for more details.

[120] James K. Loebbecke and John Neter, "Considerations in Choosing Sampling Procedures in Auditing," Journal of Accounting Research Supplement, 1975: Studies in Statistical Methodology in Auditing, pp. 38-

that the auditor is rarely certain what kind of audit environment he is working in and hence would rarely be sure of which classical estimator to use.[121] Instead, they contended DUS methods can be reliably and efficiently used in all environments. They did not believe their argument was undermined by the results of the Neter and Loebbecke study because they felt that an overly conservative version of the DUS procedure was used (which the Reneau Study confirmed) and hence the DUS did not appear as favorable as it should have.

There are other reasons for the reluctance to apply DUS technique to all sampling situations. Kaplan has identified two theoretical objections raised to the use of DUS in moderate to high error rate populations. One is that DUS is not as well designed to cope with understatements as it is to overstatements existing in the population (although this has been somewhat resolved by Vanecek's work). The second, is the unstated α risk (rejecting essentially correct populations) associated with using the more common versions of DUS. Kaplan in fact has shown that when both the α and β risks are explictly controlled for, the sample sizes are much larger than had been previously suggested.[122]

52.

[121] Anderson and Leslie, pp. 56-57.

[122] Robert S. Kaplan, "Sample Size Computations for Dollar-Unit Sampling," Journal of Accounting Research Supplement 1975: Studies on Statistical Methodology in Auditing, pp. 126-133. Kaplan assumes that DUS in practice always computes a discovery sample size. That this is not always the case, especially in Canada, is evident in reviewing the pertinent audit manuals. Also, Teitlebaum and Robinson have made a

The upshot of all this is, of course, that the classical versus DUS sampling methods controversy is still unsettled; and, even more basically, the general philosophy of whether or not to use one method for all audit environments is unsettled. The point of view taken in this dissertation will be that it is important to search for one general purpose sampling emthod. The reasons for this assumption include the persuasive arguments Anderson and Leslie have put forth concerning auditor uncertainty about the sort of errors a population may contain. It appears that to adopt the Neter and Loebbecke approach, one has to assume auditor knowledge about a population which, in fact, he is trying to ascertain by sampling in the first place. And if it is assumed the auditor does not have reliable information about the environment, then which method he initially uses should be based on how well the method works in general.

Another reason the general purpose sampling approach will be the one used is because, referring back to the Neter and Loebbecke Study, although no single method has been found to predominate in all situations, the stratified mean-per-unit estimator was reasonably effective in all of them.[123] If it were thus necessary to select an all purpose

---

comment in this regard (pp. 96-97) of their paper referenced in footnote 112). Finally, it should be noted that the planned α risk Kaplan controls for in DUS (some minimum amount of error) is different from the planned α risk normally used in classical sampling (zero errors, see footnote 80), thus the sample sizes computed for these two sampling methods are not comparable because they control for different kinds of α risk.

[123]See figure 11.2, p. 138 of the Neter-Loebbecke Study.

classical variables estimator, the most logical choice, based on the available evidence, appears to be the stratified mean-per-unit estimator.

The Neter and Loebbecke Study has probably become the most important and frequently referenced empirical research in current audit sampling literature. For this reason, it will be briefly reviewed here to put a perspective on its impact. The study used the simulation methodology to analyze the behavior of statistical estimators in sampling accounting populations. Several study populations with a variety of error rates were created from four actual accounting populations. A large number of samples from each population for each sampling plan were taken and the estimates obtained from these samples were analyzed for behavior characteristics. It was found that "at least four environmental characteristics affect the performance of a sampling procedure: 1) skewness of population, 2) error rate, 3) magnitude of errors, and 4) direction of errors. These four factors affect: 1) the precision of the estimator and 2) the reliability with which the nominal confidence coefficient (usually based on the assumption of a normal distribution) indicates the actual probability that the procedure provides correct confidence intervals.[124] It was found that most sampling plans are subject to significant unstated risks of changes in standard

[124]Neter and Loebbecke, "Choosing Statistical Sampling Procedures," pp. 41-42. Considering the results of the Neter-Loebbecke Study, Beck's work, and the issues raised about DUS, it appears that the single most important environmental factor affecting the performance of a statistical estimator in auditing is the error rate.

error and/or low reliability (true confidence less than stated confidence).

In addition to the general weaknesses discussed earlier, the Neter-Loebbecke Study failed to integrate the other parts of the audit process into the modeling effort. That is, no adjustment to planned α and β risks was attempted for the extent of internal control reliance. Constant sample sizes of either 100 or 200 items were taken for each sampling method. This is certainly not very representative of many audit situations when statistical sampling is used. The auditor frequently sets his reliability level and precision based on his concept of materiality, internal control evaluation, and analytical review. Thus Neter and Loebbecke's use of the same sample sizes in both low and high error conditions can be misrepresentative of many situations. As a consequence, their estimates of the true reliability under actual audit conditions of a sampling method may be invalid.[125] Like most of the earlier studies, the model is incomplete because it considers only the substantive testing stage of the audit process. There also exists the usual limitation of the simulation approach that only a finite num-number of situations can be analyzed. In this particular case, the critical limitation is the kind of accounting populations considered.

---

[125]In the dissertation, using the procedures that appear to be used in practice, sample sizes for stratified mean-per-unit method range from 60 to 237, and for DUS the range is 28 to 120--these are considerably different from sample sizes used by Neter-Loebbecke and, for that matter, in any of the other studies.

However, there is no indication in the literature that their populations are not considered representative of those encountered in practice.

A more formal survey of accounting environments has recently been completed by Ramage, Krieger, and Spero.[126] The analysis involved 97 audit populations of which accounts receivable book value distributions predominated (65 out of 97). The distributional forms of book values and errors have not yet been presented but some attribute and variables characteristics have. The available information indicates that most accounting populations have characteristics similar to the assumptions made in the simulation (see chapter three). For example, most monetary error rates for accounts receivable in the survey (at least 56 out of 65) fall within the error rates considered by the simulation. The study found that the error rates did not vary significantly with the size of the book value and that the size of a given monetary error tended to be proportional to the book value. In addition overstatement errors tend to be far more numerous than understatement errors (for asset accounts).[127] Finally, the study considered the proportion of what it called contamination errors in the audit population. These errors are defined to be errors such that $|\frac{y-x}{x}| > 1$,

[126] J.G. Ramage, A.M. Krieger, and L.L. Spero, "An Empirical Study of Error Characteristics in Audit Populations," paper presented at a University of Chicago symposium on empirical research in auditing, May 1979.

[127] Ibid., p. 11

where y = book value and x = audit value. (Note this is not the same as the tainting definition used in DUS. See p. 95.) The proportion of such errors out of all errors tended to have a mean of about .5 over all accounts receivable populations.[128] The simulated environments used in the dissertation have properties consistent with all of these characteristics (see chapters three and five).

2.6 Synthesis

In summary, a review of the statistical sampling literature in auditing indicates there are two competing general purpose substantive test procedures in auditing. However, the research methodology previously used does not allow any definitive statements to be made about the relative performance of the two procedures. The findings still do not allow a conclusion to be reached about the validity of the, as Kaplan so colorfully put it, "no free lunch" hypothesis:

> "With its exclusive concern for detecting errors, even under worst case assumptions, the "no free lunch" hypothesis leads one to suspect that the DUS procedure must be sacrificing something somewhere else--either in sample size or in α-risk. In my paper delivered at the conference, I tried to explore the sample size implications of simultaneously attempting to control the α and β risks in DUS. The conclusions of that paper are that if we wish not to reject populations which have trivial but nonzero error rates in them, we may need larger sample sizes in DUS than had previously been indicated. Again, this is a question on which additional empirical evidence would be beneficial."[129]

---

[128] Ibid., p. 16, averaging across their table 16.

[129] Kaplan, "Synthesis," Journal of Accounting Research Supplement 1975 Studies in Statistical Methodology in Auditing, pp. 141-142. Als Also see footnote 133.

This deficiency of prior research reflects a more serious and subtle weakness, and that is the failure to obtain direct measures of the actual α and β risks of the statistical estimators. One expects that these risks (particularly the α risk) can vary considerably depending on the amount of error in the population. The available evidence does not allow the prediction and hence the comparison of these risks for various sample sizes, associated confidence parameters, and error conditions of an accounting population. This is a serious matter for the statistical validity of the internal control hypothesis because there is no assurance then that with different sample sizes and confidence levels (as a result of internal control information) the actual sampling risks are being held unchanged. The fact that many versions of DUS and classical sampling are unreliable for typical auditing conditions makes it risky to extrapolate the evidence available to new conditions. Considering that under more realistic auditing conditions (when internal control information is used) both sample size and confidence level can vary significantly from the levels used in prior research, the statistical validity of the internal control hypothesis is an open question. This is particularly true when one considers the many linkage rules that are available, each potentially affecting substantive sample planning in significantly different ways, but none of which have been tested under realistic conditions.

Considering that the prior research on the performance of statistical estimators in auditing has not provided sufficient evidence on the statistical validity of the internal control hypothesis and that prior behavioral research indicates many auditors are not behaving in conformity to the hypothesis, it appears that statistical research should be directed to testing the hypothesis. In addition, research on internal control models has indicated that the data and analytical computational requirements of many of the models can grow to hugh proportions. Yet there is little research evidence available to guide auditors in assessing the value of such information. This is because the value is very much a function of the validity of the internal control hypothesis, and this is largely unexplored research issue.

In analyzing the ramifications of the internal control hypothesis, it becomes apparent that certain basic assumptions about the characterization of the audit process must be considered. In fact, the internal control hypothesis in general is inseparable from the linkage issue or the substantive testing method since the value of internal control information depends heavily on the performance and efficiency of the substantive testing method, and on the kind of linkage procedure used to determine the extent of substantive testing. There are many possible combinations of linkages and substantive testing methods and many appear to be used in practice. This means internal control information may have different values for different auditors. In short, it can be misleading to treat the issues of value of internal control information, performance of linkage rule, and performance of substantive testing method independently because they are to a great extent interdependent in an actual audit situation. Thus, attempting to form conclusions and recommendations to auditors on the basis of research on just one stage of the audit process may lead to serious errors.

What appears to be needed now is research to directly test the validity of the internal control hypothesis by using an analysis complete and integrative enough to parallel the usual audit process. The next two chapters describe a simulation study which is, hopefully, sufficiently comprehensive to yield more valid results concerning statistical sampling methods, linkage rules, and value of internal control information.

CHAPTER THREE

Simulation of the Accounting Environments

3.1 Introduction to the chapter

This chapter describes the part of the simulation dealing with the generation of accounting environments. The environments are described via the records making up the files which represent the environments. The assumptions underlying the environmental relationships are specified and supported. Issues relating the implications of the assumptions for the generality of the simulation results are explored. Due to the complexities of some of the issues generated in modeling an accounting system, part of the analysis is transferred to appendixes I and II. Appendix I deals with the validity of the degree of abstraction represented by the files described in the chapter. Appendix II deals with the justification of the conservatism of assuming all attribute rates (compliance error rates) in the system are equal.

The next section of the chapter describes the record format of the files which define the environments. After that follows a section defining the specialized terms that are used throughout the dissertation. These terms are also included in the glossary which is listed as appendix III. Next is a section which discusses the major assumptions underlying the relationships used in the simulation of the environments, followed by a description of the general characteristics of each environment. Finally, the chapter considers the implications of the assumptions for the validity of the simulation results.

3.2 Description of the records composing the environments

There are five accounting files, each file representing the results of a different state of internal controls. Each file is composed of records having the following general data items or fields.

RECORD I.D.	BOOK VALUE	AUDIT VALUE	K_1	K_2	K_3	K_4	K_5

where:

AUDIT VALUE is the value the auditor would agree should have been recorded but not necessarily the amount actually recorded.

BOOK VALUE is the amount actually recorded by the accounting system for that record.

K_1, K_2, K_3, K_4, and K_5 are attributes possessed by the accounting system which processes the records. All attributes are dichotomous, that is they either exist or fail to exist for a particular record. The attributes represent the success or failure of the execution of a processing task or an internal control procedure for each record. A zero value indicates the procedure was successfully completed for that record and a one value indicates unsuccessful performance.

In order to give a concrete example of the kind of internal control system that can be represented, one can imagine that the five K attributes are described as follows:

K_1 is the attribute that the amount and other data in the subsidiary ledger agree with the sales journal entry.

K_2 is the attribute that the amount and other data in the sales invoice agree with the sales journal entry.

K_3 is the attribute that quantity and other data on the bill of lading agree with the duplicate sales invoice.

K_4 is the attribute that pricing, extensions, and footings are correct and initialed.

K_5 is the attribute that quantity and other data on the sales order agree with the duplicate sales invoice.[1]

These attributes would typically be a subset of attributes the auditor would look for in the billing and shipping functions. The sampling unit in this case would be the duplicate sales invoice. However, this is an example interpretation only. There is nothing to prevent the reader from interpreting the attributes differently or from even imagining a different system. The model is perfectly general in that respect.

To show this, consider the internal controls that are relevant for assuring the accuracy of accounts receivable. For such a highly aggregated financial statement item as accounts receivable, the previous example interpretations of the K values may be at the first aggregation level only. That is, the above set of controls for the billing and shipping functions represent only one transaction subsystem impacting

[1]These interpretations are taken from a case illustration in Arens and Loebbecke, pp. 299-308, in particular p. 306.

on the accuracy of the summary outstanding accounts receivable. Other subsystems would be those relating to, for example, cash receipts, sales returns, charge off of uncollectible accounts, and allowance for doubtful accounts -- each of which in turn is affected by a separate set of internal controls.[2] It can also be imagined that if each such sybsystem were evaluated (for example as a result of applying reliability theory to subsystem component data or as a result of a direct test of the subsystem itself) into either some aggregate compliance error rate or, equivalently, a subsystem monetary error rate; then each K value could be interpreted as the net impact of the associated subsystem on the particular outstanding balance in the record at year end. Thus the level of aggregation of each attribute can be left open so that the record representation is perfectly general in that respect.

Similarly, since the book and audit values can be from any part of the financial statements, this basic record format can represent any account and its related internal control system. Thus the record format can be used to represent any financial statement items for any level of aggregation of internal control system data (as long as the internal control data is of the binomial form implying that the reliability approach is being followed).[3]

---

[2]In fact auditing standards appear to stress evaluation of internal controls by different classes of transactions as indicated here because "Controls and weaknesses affecting different classes and transactions are not offsetting in their effect." See AICPA, Auditing Standards, Sec.320.67. This paragraph also implies that each such internal control subsystem should be treated independent of the others.

[3]It turns out for many substantive tests concerning aggregate account balances, internal control information can consist of not only

RECORD I.D. is the identifying number of the record such as an account number. Since this number uniquely identifies each record in the file, this is the number by which the computer accesses the record in the file.

---

attribute data but variables data (in the form of dollar errors) as well. This arises because many tests of transactions supporting the account balances are dual purpose tests. That is, the tests make a record not only of the compliance deviations encountered but any associated monetary errors from the transaction as well. These tests of transactions are usually considered to be part of the testing of the internal controls. For example, see Arens and Loebbecke, pp. 195. Thus internal control system information for account balances consists of both attribute and variables data generally and this would appear to argue for using the Grimlund approach to modeling internal controls. Especially so since "in virtually all tests of transactions, some substantive testing is performed simultaneously with compliance testing". See Arens and Loebbecke, pp. 195.

However, the key word here is "some". That is, typically, dual purpose tests are performed on only a small subset of the sample data for which compliance testing is done and this is primarily to confirm that the control is operating as intended. "For example, assume an accounting clerk stamps a vendor's invoice after he has tested the document for clerical accuracy, proper classification in the purchases journal, and consistency with supporting documentation. A compliance test is performed by examining the invoice for the initials of the accounting clerk, and a substantive test is accomplished by actually performing the same procedures that were done by the clerk to determine if monetary errors exist. If the clerk's initials are on all invoices, and the auditor believes the clerk is independent and competent, the substantive tests can be greatly reduced but they cannot be eliminated." Thus a test of transactions is primarily a test of compliance (i.e., an attribute test) and only secondarily a substantive test (particularly when the control is not relied upon). See Fig. 6-1, pp. 197 of Arens and Loebbecke. Also Arens and Loebbecke on their pp. xiv state: "Attributes sampling is used more extensively in auditing practice than variables sampling...".

The only explanation the researcher has found for putting so much emphasis on attribute data in tests of transactions is that since the compliance error rate is typically higher than the associated rate at which monetary errors occur (this is even recognized in AICPA, <u>Auditing Standards</u>, Sec. 320B.19), it is possible to construct a smaller sample based on discovering compliance deviations (see Anderson pp. 322, especially his note (11)) than it is based on discovering the less numerous

Each file contains 7026 records having book values obtained from the trade accounts receivable file of a medium-sized manufacturer.[4] This particular book value population was chosen for the following reasons: (a) it appears representative of what auditors encounter in practice; and (b) it has a somewhat higher skewness than average for accounting populations (second highest skewness of the four Neter-Loebbecke populations) and should therefore provide a reasonable test of the appropriateness of using classical estimators for accounting

---

actual monetary errors. An interesting smoke-fire analogy has been developed by dollar-unit sampling advocates that reasons it is cheaper to design a test to detect smoke (compliance deviations) than it is to detect the underlying fire (monetary errors) directly, yet obtain the same result (a fire alarm). This implicitly assumes that in the real world there is a strong relationship between the compliance deviation rate and the amount of total monetary error in the system so that the signals are comparable. Note that with this reasoning auditors are primarily concerned with obtaining a reliable signal about the probability of material errors and that a smoke detector (compliance testing) is cost/benefit superior to a heat detector (substantive tests) in most tests of transactions. However, due to the importance of direct tests of balance sheet accounts, substantive tests are the primary purpose of tests of such accounts. See Arens and Loebbecke, pp. 197.

The ramification of this practice on internal control modeling is that although a Grimlund approach may be conceptually superior, in actual present practice there would probably be insufficient data from which to model transaction error size processes (in a very reliable system, which is often the case, monetary errors would be very rare considering that samples are constructed mainly for detecting the more numerous compliance deviations discussed above), and to this would have to be added the problem of integrating tests of transactions which are not dual purpose (i.e., attribute data only). For example, the auditors from Clarkeson, Gordon, & Co. appear to feel the sample sizes which are designed to find acceptable more than two compliance deviations are generally uneconomical (Anderson, pp. 324). This discussion thus further justifies use of the reliability approach, particularly at the account balance level, on the basis of the data availability issues discussed in pp. 70-71 of chapter two.

[4]Population Three from the Neter and Loebbecke Study, pp. 21-25.

populations.

Note all five files have the same book values and the same account numbers in common. What distinguishes the files are the audit values and the rates associated with each of the five attributes (K_1, K_2, K_3, K_4, and K_5).

The reason five attributes have been chosen to represent the internal control system is that it appears to be a reasonable number to choose to illustrate the complexities of integrating internal control data. At the highest level of aggregation of attribute data, five or six subsystems or transactions appears to be the maximum that affect a single account. For example, accounts receivable usually has five basic transactions which can affect the outstanding balance: These are the entries to sales, cash, bad debts expense, sales returns and allowances, and allowance for uncollectible accounts.[5] If each of these is thought of as an independent subsystem of accounting controls, and in practice they should be independent, then they represent five major sources of error from the accounting system. Of course the total set of controls and sttributes may be much larger as illustrated just by the attributes associated with billing and shipping function which is part of the sales recording process; but it appears that five attributes is sufficient for illustrating the complexities that arise in the integration of internal control information. This is particularly true when one realizes that to essentially test the validity of the internal

[5]See Arens and Loebbecke pp. 231-236, entries to accounts receivable are part of the larger system that is referred to in auditing as the sales and collection cycle.

control hypothesis only one attribute would suffice.[6] Arguments against considering five attributes are the following:

1. Optimum bounds on system reliability are not obtainable[7]

2. A more complex system is not essential to addressing the basic issue of the validity of the internal control hypothesis

3. The simulation is made more complex by using five attributes instead of one--this includes the need to make an assumption about the relationship of the error rates

Since the use of five attributes appears to lead to considerably richer results that can be obtained from the study at not too great an increase in additional complexity, five attributes are used in the simulation.

3.3 Key definitions

Assuming that more than one attribute applies to the internal control system necessitates making an assumption about the relationship of error rates for the attributes as well as the impact of error rates on the dollar accuracy of the records. For clarity in the following discussion, the following terms are defined:

1. Compliance error rate is the proportion of records or dollars in the file having a compliance error (e.g., K_1, K_2, etc.)

2. Monetary error rate is the proportion of records or dollars in the file with monetary errors (i.e., the proportion of records having a nonzero difference between the audit value and the book value)

---

[6]In fact the calculation of system reliability essentially involves a reduction of all the attribute data about the various subsystems to a single attribute concerning the overall system.

[7]See footnote 55 of chapter two.

3. Dollar error rate is the equivalent proportion of dollars in the file completely in error. Let A = the total of all the audit values in the file and let B = the total of all the book values in the file, then net dollar error rate can be concisely defined as $\frac{B-A}{B}$

4. Monetary error for a record is book value for a record - audit value for a record

5. Total dollar error is B - A = dollar error rate x B. Note that if total dollar error > 0 then there is a net overstatement recorded in the books and that if total dollar error < 0 then there is a net understatement recorded in the books.

It should also be noted that most of the audit literature has been rather vague in distinguishing between 2. and 3. above even though they have important implications for the way an accounting environment is modeled.

3.4 Assumptions underlying the simulation of accounting environments

With the above definitions it is now possible to be precise in discussing issues pertaining to the files.

Since very little is known about actual error rates in the business environment, it is necessary to make some rather arbitrary assumptions about these error rates.

Assumption #1. It is assumed that the compliance error rates are equal. Essentially this is an arbitrary assumption because there is no evidence about the relationships for error rates of internal control points. However, if the auditor were to assume this in designing his compliance tests, it would result in a conservative assessment of internal control reliance and hence reduce the risk of unwarranted reliance. This is shown in appendix II. Since unwarranted reliance increases the risk of making a Type II error, a conservative assessment appears to be preferable, especially since the benefits derived from more accurate assessments are at present nebulous (which in fact provides the main justification for doing the present study). Making this assumption also reduces the complexity of the simulation without detracting from its validity.

One might argue that the organizational climate would tend to limit error rates to particular ranges because of common factors such as the quality of internal auditing. Hence wide differences among compliance error rates should not be expected. In fact this argument may be the implied justification for having the auditor limit testing to only a part of the internal control system such as for critical compliance deviations.[8]

---

[8]On any audit a certain amount of abstraction takes place when the auditor must decide what are the relevant attributes of the internal control system and the types of errors which can arise and must be controlled for. That is, the auditor must identify the controls on which he plans to rely so that he can test them for compliance. This abstraction has given rise to the concept of what some auditors call critical compliance deviation. "A critical compliance deviation is a condition observed in a particular sample item which evidences a departure from a <u>key</u> control procedure on which the auditor had wished to place considerable reliance. Normally a key control is one in whose absence the occurence of a non-trivial monetary error rate would, in the auditor's judgment not have been caught elsewhere in the control system." See Anderson, pp. 323.

Assumption #2. It is assumed that the internal control system is organized as a series system. This is a less arbitrary assumption than the first one because reliability theory shows that any system of components or controls can be reduced to a series system or parallel system representation (but not necessarily of exactly five components).[9] The series system representation is the more important one because it is the one frequently used to obtain a lower bound estimate of system reliability (the upper bound estimate, for which the parallel assumption is useful, is generally of considerably less importance). The lower bound estimate is more important because it gives the minimal reliability that can be placed on the system at a specified level of confidence. To reduce the risk of unwarranted reliance the auditor would concentrate (as do engineers) his efforts on assessing the lower bound of system reliability. Hence a series system assumption is the most natural to make given that the lower bound is the one of more interest. This is synonymous with saying that the auditor wants to have a certain level of assurance that the system error rate is not above a certain amount (i.e., lower bound on system reliability = 1 - upper bound on system error rate).

Perhaps the most important reason for making a series system assumption, however, is that it is consistent with the concept of critical compliance deviation. That is, the probability of the correct processing of an account must equal the product (assuming independence

---

[9]More precisely, a coherent structure is implied, which is the usual type, and for these types of structures it is possible to represent the system as a minimal path series structure or a minimal

of the controls) of the probabilities that each key control procedure will not result in an error. These probabilities are the reliabilities, Ri, associated with each key control subsystem i, and hence the probability that the whole system results in correct processing (i.e., no errors in the account) is the product $R=R_1 \cdot R_2 \cdots R_n$ where n is the number of key controls in the system. This is the same as the reliability of the system assuming it is serial.[10]

Assumption #3. It is assumed that the five internal control points or attributes are independent. This assumption is made to make the mathematics of estimating system reliability more tractable.[11] This is another assumption which is less arbitrary than the first because the qualities of a good internal control system such as separation

---

parallel cut structure. See Barlow and Proschan, pp. 6-11.

[10]It should be noted that the amount of errors is conceptually additive using the Grimlund approach, but that in the reliability approach, modeling of the error size is not done except at the aggregate systems level. Thus using the reliability approach assumed for this simulation for the reasons given on pp 71-5 of chapter two requires only aggregation of error rates at the subsystems level. This means using the multiplication rule given because any number of critical compliance deviations beyond one (e.g., not all K values equal to zero) for an account, does not change the fact that the account has a compliance error associated with it. This is the relationship the multiplicative rule for series systems reflects and not that any accompanying monetary errors are additive (e.g., a monetary error in recording a credit sale is not necessarily offset by a monetary error occurring on recording cash receipts).

The systems reliability formula for a series system can be found in David K. Lloyd and Myron Lipow, Reliability: Management Methods and Mathematics (Published by the authors, Redondo Beach, California 2nd Edition) pp. 222.

[11]All of the models reviewed in this dissertation, including Grimlund's most general model, requires the assumption of statistical independence between separate accounting functions or subsystems. For

of duties appear to encourage independence of critical internal control procedures.[12] Also this is a common assumption made by engineers in their application of reliability theory and it appears to work well for their purposes.[13]

A point which should be made in this regard is that if the dependency among internal control components is of a particular form then the series system assumption above still insures a lower bound estimate of system reliability. The form of dependency for which this is true is where the internal controls are "associated". Association is a generalized form of positive correlation between two random variables. Thus if the internal control system is such that if one fails the probabilities increase that the others fail, the series system assumption still provides a lower bound of the system reliability estimate.[14]

---

example see Grimlund "A Framework for the Integration of Auditing Evidence," pp. 47-49 for a fairly complete discussion of this issue. Generally, without such an assumption the mathematics is even less developed than with it. Fortunately, this appears to be a valid assumption to make for most internal control systems.

[12]In fact by the very definition of key internal control procedures, independence is implied. See footnote 8. Also see Arens and Loebbecke, pp. 194, and the AICPA Auditing Standards, Sec. 320.67,320.36, and 320.59 for support that internal controls should be independent.

[13]The reasons for this should perhaps be recorded because they wound so familiar to the auditor's plight. "For years we have been relying on simply series models for large systems with complex structures. The reasons generally given for this oversimplification are ignorance of the detailed reliability structure and a dearth of component data. Both these factors are slowly changing, and engineers are beginning to use more complex system structural models. The computation involved in such cases becomes laborious, and use of a digital computer is generally required." From Martin L. Shooman, Probabilistic Reliability: An Engineering Approach (McGraw Hill, New York, 1968), pp. 210. Also see pp. 120 and pp. 213.

[14]Barlow and Proschan, pp. 29-32.

At this point it appears appropriate to summarize relevant aspects of engineering literature to see how assumptions 2. and 3. fit into the auditing framework. The auditor's perspective is the following. He is in a position to test the output of the system directly (the output of the system is the dollar accuracy of the records) but he attempts to reduce the amount of testing of system output via knowledge of the performance of the components of the system (as well as other factors such as system design). This has apparently come about because testing of components is frequently cheaper than testing the output of the system and because much component testing (in the form of tests of transactions) is required for the auditor to obtain evidence on the fairness of presentation of details within the financial statements (e.g., income statement items).

This auditing philosophy is difficult to criticize. It appears reasonable that the auditor should be able to reduce the amount of testing of system output (e.g., account balance tests with variables estimators) if he has convincing evidence that the components of the system are reliable. The problem arises when it must be decided how much to reduce such system testing. This is undoubtedly a function of the accuracy of the system modeling effort and the availability of data about the system components.

The way auditors presently use component (i.e., internal control) data for reducing substantive tests is in controlling the risk of unwarranted reliance. That is, the auditor relies on the internal controls to reduce system (substantive) tests when there exists no material compliance error rates (and hence, presumably, no material total dollar error). So the auditor is typically very concerned that at a specified level of assurance (confidence level), the compliance error rate is not above a certain amount.[15] Hence the upper error limit is of paramount concern and the decision on reliance is generally based upon it.[16] However, this is statistically equivalent to stating that the internal control procedure has a lower bound of reliability equal to that same level of assurance. This relationship holds because by definition reliability = 1 - error rate, for a particular internal

---

[15]This certain amount is the material amount and requires the auditor to establish a relationship between compliance error rates and the amount of monetary errors. This in turn requires some modeling mechanism such as one based primarily on subjective assessment (i.e., the traditional approach) or more explicit approaches such as the reliability and Grimlund approaches discussed in chapter two.

[16]Support for this statement can be found in AICPA, Auditing Standards, Sec. 320B.22; Arens and Loebbecke, pp. 286 who state "A one-sided interval generally specifies an upper bound only and represents the probable worst likely error rate. This type of attributes estimate is the one most commonly used in tests of transactions." And in a recent paper devoted to this topic the indication is that the upper error limit is the primary criterion for the decision on the degree of reliance: "We believe attribute sampling to be the most common form of compliance testing at the present time when statistical sampling is used. As referred to earlier in this paper, upper precision limits [same as upper error limit] are usually selected between 1 and 5% and in some cases up to 10%. Reliability levels [confidence level] usually range from 90-95%." From William L. Felix, Jr. and James L. Goodfellow, "Audit Tests for Internal Control Reliance," paper presented at the American Accounting Association Convention, August, 1978.

control procedure.[17]

It is therefore apparent that in using the reliability approach the auditor's interest on internal control reliability will center on a lower bound estimate much as it does for engineers and for very similar reasons (to reduce unwarranted reliance).[18]

It is now possible to discuss the significance of making a series assumption for the system of five internal controls considered in the simulation.

First consider the case where component reliabilities (R_1, R_2, R_3, R_4, R_5) are known for certain for a system of five independent components. Then it can be proven that no matter how these components are organized, a lower bound of system reliability is given by the product $R_1 \cdot R_2 \cdot R_3 \cdot R_4 \cdot R_5$, i.e., assuming the system of components is organized as a series system. (Similarly a guaranteed upper bound can be computed by assuming the system is organized in parallel, i.e., an upper bound is $1-(1-R_1)(1-R_2)(1-R_3)(1-R_4)(1-R_5)$.)[19] Also note this is guaranteed even if the reliabilities are dependent by being associated.[20]

---

[17]Here error rate is left purposely vague (it can be either a monetary error rate or compliance error rate) so that the concepts can be discussed in full generality. Either way lower bound on system reliability = 1 - upper bound on system error rate.

[18]And most importantly this is completely statistically equivalent to the traditional approach which makes interpretations on upper error limits. See footnote 16. Also see Mann, Schafer, and Singpurwalla, pp. 373.

[19]Barlow and Proschan, pp. 33.

[20]Ibid., pp. 33.

This is what is meant by saying that a series assumption is conservative for any given set of components. Of course, if the system is series organized then a series assumption yields exact system reliability.[21] On the other hand, if someone were not sure how the components were related, then a series assumption assures they will not overestimate system reliability (and this is an important consideration if unwarranted reliance is to be contained within desired bounds).

The series assumption is significant for another reason. This is that for the kinds of systems encountered in the real world, it is possible to identify certain subsets of the components of the original system (these subsets are called minimal cuts, minimal sets, and modules depending on what structure they take) and by working with the reliabilities associated with these subsets to obtain a series representation which is then used to compute a lower bound.[22] These lower bounds can sometimes result in considerable improvement over the crude bound obtained assuming every component is connected in series to every other component. On the other hand, if every component happens

---

[21]As noted earlier, this is very likely the most realistic structure assumption to make in modeling internal controls. See footnote 8 and pp. 123-125 of chapter three.

[22]There are many ways to compute exact system reliability (assuming component reliabilities are known with certainty) for a general system with a coherent structure. (See Barlow and Proschan pp. 24-25 and Shooman, pp. 129-140.) However, for complex systems this can be a very formidable task, so simpler methods for computing bounds on system reliability have been developed (for example, see Barlow and Proschan chapter two and appendix, or Shooman, pp. 202-264). Thus an argument for making a series assumption in auditing (when the controls are not actually series structured) is that such an assumption leads to a conservative reliance which is consistent with the general conservatism

to be organized in parallel, then the series representation is that where the entire system is treated as a single component with the single component being the system itself.[23]

The implication of this fact for the simulation is that, again, each attribute can be assumed to be associated not only with a single component or control point but with an entire sybsystem as well. The K values represent the decomposition to some level of aggregation (or even varying levels of aggregation) for which binomial data (i.e., pass/fail data) is available as a result of an internal control examination.

A series assumption reflects the well known principle that "a chain is as strong as its weakest link". And it appears that this idea is relevant for certain factors that auditors consider about an organization: probability of management override, system design evaluation, and compliance with the system. Up to this point reliability has been discussed implicitly within the framework of compliance with the system. That is, if certain procedures and controls have been set up, are they properly being complied with and are they operating satisfactorily. It turns out that the other two factors, system design and management override potential, appear to be best represented as a series

---

of audit practice and accounting theory.

The methods that are being considered here are the ones computing a lower bound since as argued earlier this is the most relevant bound when using the reliability approach.

[23]A parallel system is essentially one which is redundant. In an accounting control setting it would typically be represented by a reprocessing of a record to insure its accuracy. It appears such a control is best measured by the output of the two operations (the original processing and the check processing) so that a direct estimate of the

system with compliance. That is, material errors could occur if management does override the system even though system design and compliance may be perfect. Similarly, material errors could occur if system design is very poor but there is no management override and there is perfect compliance. Also, if compliance is poor material errors are possible even if there is no management override and the design is perfect. It is thus apparent that the auditor cannot reduce his substantive tests if any of these factors are considered to have weaknesses. Each is a potential source of error which cannot be corrected or detected by the other two. Essentially then these factors represent a series system in terms of error generation in the records.[24]

So far in this chapter the discussion about system reliabilities has assumed perfect knowledge about component reliabilities, even for the computation of the bounds on system reliability. In fact component reliabilities are but estimate themselves. Hence the system reliability estimation problem for a general system is even more complicated with estimated component reliabilities. At present this is a largely unsolved problem. However, the techniques available for estimating the

---

resultant subsystem error rate is obtained (see footnote 3 for an example). In general the relationships among subsystems of controls can be expected to be serial. See pp. 123-124 of chapter three.

[24]Kinney, pp. 23, also hypothesized a multiplicative rule for system design and compliance but like Warren, pp. 11, had doubts about the weighting that might be implied. To this researcher, attaching equal weight to these factors and the series multiplicative rule seems most natural (being consistent with elementary probability theory) and defensible.

lower bound of system reliability for complex systems from estimates of component reliabilities generally require that the system be expressible as a series or parallel system composed of additional series or parallel system.[25]

Assumption #4. The dollar amount of error for any particular record should be held between plus or minus 100% of book value. It is necessary to set a specified limit because DUS needs to make an assumption about the site of the worst undetected error. According to Teitlebaum, who has worked closely with the firm of Clarkeson, Gordon, & Co., "the risks of a tainting exceeding one are slight."[26]

Assumption #5. Errors of monetary understatement as well as overstatement should be generated.[27]

Assumption #6. The size of dollar errors should tend to be proportional to the size of the book value of the record. Support for this assumption is implied by the common audit practice of automatically

---

[25]See Mann, Schafer, and Singpurwalla, pp. 518.

[26]Teitlebaum, "Dollar-Unit Sampling in Auditing," pp. 19 of appendix III. Also see the Clarkeson, Gordon, & Co. manual, pp. 125 and Anderson, pp. 351 for additional evidence. The net result of this evidence is that at least for asset accounts the likelihood of greater than 100% errors appears small.

[27]Teitlebaum, pp. 27 states, "most fields will contain errors of overstatement and understatement."

examining accounts over a certain size.[28]

Assumption #7. Total amount of dollar error in the file can be assumed to result in a net overstatement of the total audit value amount for each accounting environment simulated. In the situation of a file of records for assets, this is assumed to result in the case of most concern to auditors. That is, the general bias for errors in recording assets is assumed to be that of overstating the assets.[29]

Assumption #8. Each compliance deviation (or K value) has approximately 1/3 probability of generating a monetary error. Thus each control point or accounting subsystem has a monetary error rate equal to 1/3 the associated attribute or compliance error rate.[30]

---

[28]According to the Clarkeson, Gordon, & Co. manual, pp. 4 "It would seem that in the majority of cases internal control procedures will be such as to make the risk of misstatement in either direction approximately proportional to size of book value..." Also on pp. 88 of Don Robert's Statistical Auditing he states that is unlikely that the magnitudes of errors are unrelated to the size of the recorded amounts. In addition Arens and Loebbecke, pp. 363, state "...in accounts receivable it may be reasonable to expect the large accounts to contain the large errors and the small ones to contain the small errors, even though this need not necessarily be true." Aso see reference in footnote 19 of chapter two. Thus assuming the dollar errors are proportional to the book values appears to be a reasonable and persuasive assumption to make.

[29]Considering from recent court cases that the effect of fraudulent activities and management bias tends to overstate assets and/or understate liabilities. This appears to be the most reasonable assumption to make assuming a significant net error will arise. See Arens and Loebbecke, pp. 137-139. Kaplan justified the use of an overstatement error model for assets (and Scott thought "that the asset valuation problem is the primary one facing the auditor." pp. 315) by the following: In general the recorded value of items in asset accounts is highly likely to be overstated rather than understated bacause of the auditor's bias toward conservatism and management's bias toward reporting higher profits." See Kaplan, "Stochastic Model for Auditing," pp. 40.

[30]There is very little evidence available on the nature of the

Assumption #9. The reliability approach as described and justified in chapter two appears to be the most relevant for present audit practice.[31]

3.5 General description of the accounting environments

With the assumptions of the previous sections it is possible to complete the specification of the system. Once the book values have been recorded in the file, the internal control system (represented by the five K values) will introduce errors into the records so that the book value will not always equal the audit value. This is done by assuming an error rate for each of the five attributes and then generating dollar errors based on this rate. If the error rate for attributes

---

relationship between compliance deviations and monetary errors. What little evidence there is comes from DUS users who are the most open in disclosing assumptions based on their experiences; and so it is their views that are most influential in the creation of accounting environments.

As originally indicated in chapter two (pp 40-1) this three to one relationship is that suggested by the firm of Clarkeson, Gordon, & Co. as being reasonable for use in auditing. In real life it is expected to be conservative (i.e., less likely to rely on good internal controls) because actual ratios are probably higher (see the Clarkeson, Gordon, & Co. manual, pp. 117-119). The simulated accounting environments remove much of this conservatism because the simulated relationships correspond to what the auditors assume. However, an additional source of conservatism that practitioners introduce is by assuming that every monetary error so generated is a full 100% overstatement. (This follows from the three times materiality rule.) In this respect the simulated accounting environments are not consistent with the DUS practitioner assumptions because only the largest monetary error is a 100% overstatement. Teitlebaum, McCray, and Leslie, pp. 12 indicate most taintings are of "moderate" size.

[31]Reasons for this are given on pp. 71-5 of chapter two. Note that these reasons apply for substantive tests of account balances as well as transactions. See footnote 3 for an explanation.

(compliance error rate) is given by ϕ, and X_i represents the book value of record i; then the audit values Y_i are generated as follows:

$$Y_i = X_i \text{ with probability } \left(1 - \frac{\phi}{3}\right);$$

$Y_i = \theta X_i$ with probability $\frac{\phi}{3}$ where θ is a random variable with a beta distribution $F(\theta)$ between values 0 and 2 and with a mean between 0 and 1.[32]

The same error rate and the same error generating function $F(\theta)$ is used for all attributes. The same $F(\theta)$ is used for all files but different ϕ's are used for different files.[33]

The files are constructed so that the only way a dollar error can occur is if a compliance error occurs for a record.[34] The sequence is to first generate compliance errors. Then if a compliance error has occurred (one of the five K values is 1) in a record, a possible error

---

[32]The beta distribution (including standardized and extended forms) is fully described (including the first four central and non-central moments for both forms) in Grimlund, "A Framework for the Integration of Auditing Evidence," appendix II. The beta distribution was selected to generate the error sizes because (1) it is consistent with all the assumptions about accounting environment given in earlier sections, (2) it is relatively simple to use, (3) other researchers have suggested its use for this purpose [notably Kaplan, "A Stochastic Model for Auditing," pp. 44 and Scott, "Asset Valuation and Audit Size," pp. 317 footnote 23], (4) little is known about dollar error generating processes in accounting, and (5) it is useful for testing the robustness of Felix-Grimlund's model which assumes such errors are normally distributed.

[33]Thus the error rate is the main experimental treatment in the environments. As noted in chapter two, the error rate appears to be the most important environmental characteristic affecting the performance of statistical estimators in auditing.

[34]Certainly the impact of internal control information is largest when internal controls are the sole source of monetary errors. Otherwise, if there are significant other sources of error (e.g., management

in book value is generated by the rule given for Y_i above. If the use of the rule results in a dollar error being generated for a record (book value differs from audit value), then the computer is programmed to go on to the next account. If a dollar error is not generated for a particular compliance error, the computer is programmed to complete the inspection of the remaining five K values to search for possible other compliance errors in the record and repeat the sequence.

Note that the above procedure biases the first few attributes as being the sources of dollar errors. However, since the error generating process $F(\theta)$ is the same for all attributes, a more random allocation would produce the same distribution of errors. Hence this aspect of the programming does not result in a theoretically different distribution of errors.[35]

It should also be noted that this rule results in a situation where a compliance error does not guarantee that a dollar error is generated, rather it only increases the probability that a dollar error is generated. However, a dollar error can arise only if a compliance deviation has occurred so that the system is a closed system with respect to monetary errors.

---

override), the auditor would not want to reduce his substantive tests no matter how good the internal controls are (see the discussion on p. 131).

It is assumed that an upper bound or maximum value of internal control information is preferred for assessment because the lower bound is already known; it is zero since the auditor always has the option of not relying on internal controls. Thus a closed system appears to be the best system to use in assessing the validity of the internal control hypothesis.

[35]This really follows from the fact that the reliability approach

Use of the above rule results in a file of records such that a dollar error for a record will never be above or below 100% of book value, total book value will generally be greater than the total audit value, and the size of the dollar errors is proportional to the book value.[36] The rule also results in an internal control system which has a series system configuration. That is, the probability of a dollar error occurring in an account, using the above notation and assuming the error rate for each attribute is the same, is $1-(1-\frac{\phi}{3})^5$. (Note that this is also the monetary error rate for the file.) Alternatively, the reliability of the system is $R=(1-\frac{\phi}{3})^5$. Again, the definition of reliability used in this research, unless otherwise specified, is the probability that a record is processed without a dollar error occurring. (Note this is not the same as the probability that a record is processed without a compliance deviation occurring, $(1-\phi)^5$). This meaning of reliability is in accord with earlier accounting usage of the term.[37]

---

is being used and that $F(\theta)$ represents the summary system (as opposed to the individual subsystems -- the five K values) error size generation process. The attributes just determine when this error size generating function is called.

[36]It should be reiterated that this error pattern causes the most problems for classical statistical estimators (see footnote 135 of chapter two) in addition to likely being the most common type in practice. Thus in a sense this error size pattern represents a sort of acid test of the statistical validity of the internal control hypothesis.

[37]For example see Cushing or Bodnar.

The five accounting files produced have target reliabilities (using the definition of complement of the monetary error rate given in the preceding paragraph) of $R_1 = .99$, $R_2 = .95$, $R_3 = .91$, $R_4 = .90$, and $R_5 = .85$. The beta distribution for the error generation process was picked so that the accounting environment (E) with reliability of .90 has almost exactly the material amount of dollar error (.05 of the total book value as described in chapter one and to be further discussed in chapter four).[38] The rest of the environments have amounts of dollar error roughly proportional to the monetary error rate = (1 - reliability) for the environment.[39]

The reasons for setting these environmental values is to obtain a clear trend for the actual α and β risks incurred by an audit sampling strategy. This is made clear by referring to Figure 4 which shows the reliabilities associated with each environment.

As is evident from glancing at Figure 4, three of the environments (E1, E2, and E3) have less than a material amount of total net dollar errors and hence only a Type I error (α risk) could occur. Two of the

---

[38]An attempt has been made to justify as many of the environmental characteristics as possible. For example, the monetary error rate (1-R) associated with the environment having a material amount of dollar error is 10%. This is because Neter and Loebbecke indicate such an error rate is normally considered "high" implying it is unacceptable (Neter and Loebbecke, pp. 127). Similarly, a 5% monetary error rate is considered "moderate" and a 1% monetary error rate is considered "low".

[39]An attempt has also been made to line up the compliance error rates with the monetary error rates and dollar amount of error so that they are consistent with the few published interpretations in practice. (For example, see pp 40-1 of chapter two for interpretations of upper error limits on compliance deviations and footnote 16 of chapter three.) All actual environmental values are given in chapter five.

FIGURE 4

Sampling Risks of the Accounting Environments

E1	E2	E3	E4	E5
R=.99	R=.95	R=.91	R=.90	R=.85
			exactly material amount of total dollar error	
α risk only (i.e., immaterial error E's only)			β risk only (i.e., material error E's only)	

environments (E4 and E5) have at least a material amount of error and hence only a Type II error (combined risk) could occur.

Since R = .9 (E4) results in an exactly material amount of error, the actual combined risk is highest for this environment. Similarly, the highest actual α risk occurs in E3 (i.e., the highest immaterial total dollar error occurs with R = .91). The reason three immaterial error environments are considered is that this allows an even spread of all immateriality conditions and hence a complete range for the possible actual α risk. Only two material error environments are considered because with increased amounts of errors the combined risk will only get less than that already measured. Also, when the amount of error gets too large it is unlikely the auditor would ever attempt compliance testing and may even disclaim an opinion on the financial statement. Hence using internal control information appears feasible only when the amount of total error is not much greater than materiality, certainly not several times greater than materiality.

The actual environmental values and statistics used in the simulation are given in full in chapter five.

3.6 Significance of the assumptions

To conclude, it appears appropriate to consider some general issues related to simulation methodology and to some specific questions that might be raised about the "realism" of the simulated accounting system.

Any simulation will be but an abstraction of the real world and hence many factors and complexities by necessity have to be excluded from this abstraction.[40] However, the abstraction must include all the relevant features of the environment for the particular goal of the study. The controversy arises in deciding what is relevant and therefore to be included in the simulation.

At the same time one must be cautious about including too many irrelevant factors which increases simulation complexity needlessly and what's worse, may confound the effects of major interest. (See the end of appendix I for more discussion on this topic.)

To guide the following discussion, it may be wise to specify the basic goal of the proposed research: to test the validity of the internal control hypothesis of auditing within a statistical sampling framework. There are other goals but this is the main one and others

[40] As noted in chapter one, many of these arguments really emanate from the abstraction issues associated with models in general and probably can never be completely resolved to everyone's satisfaction.

are basically offshoots and extensions of it.

With this in mind the following statement is assumed true for purposes of deciding what is relevant to incorporate in the simulation: the key aspect of the study is to see how information about an internal control system impacts on the sample size for substantive tests and the risks associated with the audit process, where these effects are measured by simulation of alternative audit strategies in various environmental conditions. The key aspects of the accounting environment are the following: a population of accounts (book values) and their distribution, monetary error rate and distribution of dollar errors for the accounts, key internal control points, attribute error rates associated with key internal control points, and the relationships between attribute error rates and the amount of total dollar error in the file. (Although such things as causes of errors and system design are also important, they are presumed to be embedded in the consideration of the above factors.)

If there is agreement with the above statement (And, by the way, all the audit strategies described in chapter four do not ask for more information about the system. Hence any system providing this information is sufficient for purposes of testing the strategies and, therefore, the internal control hypothesis. Any additional complexity or "realism" would not affect the way the strategies would operate.), then it is apparent that questions, for example pertaining to whether a particular attribute applies to a particular level of aggregation of an account are not really germane to the goals of the study.[41] All that needs to be established in the simulation is that an attribute affects the accuracy of the book value in a certain way and, therefore, the attribute needs to be considered in the audit strategy. The internal control hypothesis is very general. Hence, if it does not hold true in a system where the internal controls define the amount of dollar error in the file, and the error rates are directly proportional to the error amount, then it cannot be expected to hold in many real systems. (Assuming that the same audit strategies are used.)

Similarly, such issues as the use of particular error rates and error generating functions $F(\theta)$ are also not germane to the basic goal of the study as long as they result in accounting values and errors which appear reasonable in terms of actual accounting conditions.

With this perspective it is now possible to address some specific questions that might be raised in regard to the described accounting system.

Question 1: The accounting system described applies only to the lowest level of aggregation of accounts -- line items (transactions) throughout the year. Therefore, it cannot be used to apply to account balances and so the conclusions must be limited to the transactions case.

Response to Question 1: This question is motivated by the fact that internal control information for account balances can consist not

---

[41]These are instead issues pertaining to modeling a particular accounting environment. While this is important for establishing the relationship between compliance and monetary errors in an actual audit,

only of attribute data for transactions as well as account balances, but also variables data from transaction tests. The latter situation can involve error size modeling at the subsystem level (i.e., what has been called here the Grimlund approach) and although this may be possible for some audits, it does not seem likely in general as discussed earlier.[42] This then leads to the former situation (i.e., the problem of different aggregations of attributes). As explained in appendix II, this can be mathematically justified on the same basis that the binomial distribution can be justified as an approximation to the hypergeometric (in fact, is the justification used extensively in audit practice); and on the flexibility available (in fact, recommended according to SAS#1 Sec 320B.2) in redefining and aggregating attributes to make them more pertinent for the level of aggregation of the related substantive test.[43] These issues are more thoroughly explored in appendix II.

---

it is not as important for testing the validity of the internal control hypothesis which assumes such a relationship has already been established.

[42]Much of the argument against the Grimlund approach has been based on data availability issues, see pp. 70-71 of chapter two; although notwithstanding Grimlund's work there still appears to be many potential analytical problems, see pp. 65-69 of chapter two.

[43]In fact this is the most important aspect of defining the attribute, see Arens and Loebbecke, pp. 301-302.

Question 2: A criticism that is somewhat related to Question 1 is that the system fails to recognize the fact that tests of compliance should be applied to transactions throughout the period under audit and not just at one point in time.

Response to Question 2: This question can also be settled by appealing to the closeness of the approximation of the binomial to the hypergeometric distribution where the sampling is from a large population of attributes. It should be noted that in actual practice most attribute testing is performed on an interim date and, as a result of other evidence obtained, the auditor frequently assumes that the reliance assessment made in the interim date applies to the balance sheet date as well; i.e., the frequent situation is that there has been no basic change in the system during the interim period. Note the emphasis is on the state of the system and not on particular errors. Hence, frequently, error estimates are made for attributes which will not necessarily affect the balances at year-end (but which did affect the balance at the interim period) because the assumption is that the same error rates will hold till the end of the period (in reliability theory this is known as the constant error assumption and the name frequently attached to this is process reliability).

In this sense then the simulated system represents an idealization of the accounting environment encountered by the auditor. That is, the compliance errors define the amount of error in the file and, therefore, by sampling for the attributes in the record the auditor directly obtains estimates of the error rates that are most pertinent to the account balances at year end. Hence all sources of error that may be introduced by changes in attribute rates during the interim period are eliminated. By trying to model all attributes during a particular period, including those that do not pertain to outstanding balances at year end, it is the feeling of the researcher that the simulation would be unduly complicated. See appendix II for a further discussion of similar issues.

Question 3: Since it is theoretically possible to analytically determine the total amount of dollar error for the proposed system of what use is the simulation?

Response to Question 3: This is an irrelevant point because the statement is also true for any "real" accounting system. That is with enough knowledge any real accounting system can be modeled and either analytically solved or the output obtained by a Monte Carlo simulation. The point is, very few if any auditors have the time or resources to gain such intimate knowledge about an accounting system.

Also, it may not be possible to analytically determine the total amount of error if for example a Grimlund approach were followed and different dollar error size generation processes were operating for the subsystem components. This is because Grimlund is the only one to have attempted to analytically deal with this problem, and his model is yet untested for accuracy and usefulness.

The main point of the simulation is not to analyze the system but, rather, it is to analyze the audit strategies (the set of rules that have evolved to help the auditor cope with the problems involved with attaining his goals in the accounting environments he usually encounters). Predicting the relative performance of the audit strategies does not appear to be within the realm of an analytical solution and so this is the main reason a simulation is resorted to.

This completes the description of the accounting environments. The next chapter discusses the audit sampling strategies that are applied to these environments.

CHAPTER FOUR

Description of the Audit Sampling Strategies

4.1 Introduction

Given the accounting systems described in chapter three, the next phase of the simulation involves a reproduction of various audit strategies which either have been proposed or are actually used in practice. The basic goal of an audit strategy is to reach a conclusion about the acceptability of the reported book values of the accounting population. This is done by sampling from the population of accounts and obtaining evidence from the sample of the total audit value of the population. The amount of sampling, as argued by the internal control hypothesis, is in turn determined by the amount of internal control information available and how that information is used (via "linkage" rules). Operationally, then, a given audit sampling strategy consists of a certain amount of internal control information, a linkage rule, and a particular variables sampling method for substantive testing for the records in the file.

The next section lists and justifies some assumptions underlying the formulation of the audit strategies used in the study. Following that is the section which describes and justifies the various alternatives in each of the three stages of a strategy which are used in the simulation. Then the individual strategies are listed and summarized. After that is a section describing the rules used for comparing strategy performance and other statistics collected in the study.

4.2 Assumptions underlying the formulation of audit sampling strategies

In order to make the subsequent data analysis feasible, it is necessary to make some assumptions about the conditions under which the sampling strategies operate.

Assumption 1. It is assumed that the hypothesis testing approach for the substantive test results is a reasonable one to follow in simulating a sampling strategy.[1]

Discussion: Support can be found for using either the hypothesis testing approach or the estimation approach in substantive testing. It is clear that either approach can be justified depending on the particular audit circumstances and goal of the auditor. The hypothesis testing framework has been chosen for this study because it greatly simplifies subsequent data analysis. Under this approach there is no need to make assumptions about the auditor's subsequent actions depending on the results of the test. Since the auditor views the substantive test of balances as the primary source of evidence about the accounting system, it is assumed that the statistical result of the substantive test is synonymous with that of the auditor's decision on obtaining the result (e.g., if the statistical test indicates rejection of the total book value, then it is assumed the auditor also rejects;

---

[1]Support for using the hypothesis testing approach in auditing can be found, for example, in Robert's Statistical Auditing, pp. 40, or in Elliott and Rogers.

this means the auditor would either expand testing, refuse to make an unqualified opinion, or take some other appropriate action based on the circumstances).

Assumption 2. Under the hypothesis testing approach it is necessary to make assumptions about the α risk, the β risk, and the materiality level that should be used in the statistical test. For purposes of the simulation, it is assumed the auditor's goals in this respect are $\alpha = .05$, materiality = 5% of total book value, and β will be adjusted on the auditor's evaluation of internal controls such that combined risk = .05. The null hypothesis is that there is an immaterial error in the book values.[2]

Discussion: By specifying the above parameter values an effective constraint is put on the form of the auditor's loss functions. So the objection might be raised that these values may not be valid because they do not properly reflect auditors' loss functions. One can respond to this objection by noting that in the present state of the art one is even less sure about a more specific representation of auditors' loss functions. In fact, at present, there is a controversy as to whose loss function should be used: The individual auditor in charge of the audit? The audit team that makes the decision as a group? The audit firm? The investors in the client firm? Potential investors? Society as a whole? In addition there would be disagreement not only with

[2]Auditing theory distinguishes between the null hypothesis that says there is no material error (often called the positive approach) and a null hypothesis that says there is (negative approach). See Roberts, Statistical Auditing, pp. 40-48. The negative approach predominates in compliance testing but both approaches are frequently used in substantive testing.

respect to specific parameter values of a particular loss function, but with the general form (e.g., linear or quadratic) of the loss function as well. Given that the goals of the dissertation do not include these issues, it appears appropriate to take the approach given here by specifying particular α, β, and materiality values. This assumes general agreement can be reached on the validity of such values.

That this may be possible is made evident by references to official pronouncements, audit textbooks, and firm training manuals. The fact of the matter is auditors appear to find it much easier to express loss functions indirectly through materiality levels and risks associated with the audit than by completely expressing the mathematical form of their loss functions. For example, a review of audit training manuals indicates most audit firms as a matter of firm policy use specified levels of α and combined risks in planning their sample sizes. Thus it should not be inappropriate to use a similar approach in simulating audit strategies.

The review of the auditing literature indicates that although differences exist, the materiality and risk levels tend to fall over very short ranges. Hence it appears target levels for these factors can be set fairly unarbitrarily. Although there may be some disagreement on the exact values to use in the simulation, it is evident that other relevant levels will not be far from the simulation values. Anyway, this is not so much a criticism of the methodology, as it is a criticism of choice of levels which are essentially judgmentally determined. (But there is considerably more hard evidence available on these levels than there is on the general form of the auditor's loss function.)[3]

By using the hypothesis testing approach with α = .05, combined risk = .05 and materiality level of 5% of book value, the subsequent data anlaysis is considerably simplified in terms of the impact of

---

[3]Perhaps one approach which can be used to identify classes of loss functions used in practice is to formally study planned risk and materiality levels used in practice to see how they conform to various functional forms.

The following references suggest parameter values either equal to the ones used in the simulation or ranges of values which include the values of the simulation (this is not an exhaustive list).

SAS No. 1 Sec. 320B.32 and 320B.35 use as an illustration combined risk = .05.

The latest AICPA thinking which appears to be reflected in Statistical Auditing by Don Roberts has many examples foremost of which would be the chapter devoted to an illustration of the application of statistical sampling techniques, chapter seven. Page 134 illustrates a planning table constructed on the assumption that combined risk = .05. On the same page Roberts states "...the combined risk level of approximately .05, which would be appropriate in practice provided...the non-sampling risk is small." Also see p. 135 and p. 170. He frequently uses β = .05, e.g., see pp. 41, pp. 170, or pp. 178. Materiality is defined to be .03 to .04 of book value and .025 of pre-tax income on pp. 157.

Elliott and Rogers who reflect the practice of Peat, Marwick, Mitchell & Co., recommend α = .05 (p. 49), combined risk = .05 (p. 50), and materiality = M = .05 of net income (p. 52).

Clarkeson, Gordon, & Co. of Canada suggest materiality = .05 of pre-tax income (unless income is very small or negative) on p. 145 of their manual, and combined risk = .95 to .99 on p. 160. They do not mention control of α risk.

Ernst and Ernst, Audit Sampling, (USA, 1976, 2nd Printing) Use combined risk = .05, α risk = .05 and materiality = .1 of book value in their illustration (p. 80).

various procedures used by an audit strategy. This implication is discussed in chapter one and again later in this chapter.

Since a sampling strategy consists of a combination of internal control information, linkage rule, and substantive testing method; it should be pointed out that the goal parameter values used in the simulation reflect the intended performance of the strategy as a whole and not just a substantive testing method. In particular this means that β = risk of Type II error (with null hypothesis that there is an immaterial error in the total book values) for substantive sampling is determined by the combined risk level for the strategy (risk of making a Type II error as a result of applying the entire sampling strategy and not just the substantive test), and the particular linkage rule used. These linkage rules reflect the essentially Bayesian philosophy in auditing that auditors can reduce the amount of substantive testing as a result of reliance on internal controls. This is operationalized statistically by letting the β value for the substantive test increase whenever controls are to be relied upon.[4]

The α level and materiality level, on the other hand, are the same for both the audit strategy and the substantive testing procedure. This

[4]Note that under this reasoning the auditor expects planned β to be significantly different from the risk of Type II error that he actually expects to experience as a result of applying a strategy. For example, when the auditor sets planned β at .5 for his substantive test as a result of internal control reliance, he is really still expecting his overall risk of making a Type II error to remain at .05, his planned combined risk. The way this is operationalized depends on the linkage rules which are fully described in section 4.3.

is because the impact of internal control information on an audit strategy is made by means of the β level on the substantive test using a linkage rule.[5] Thus α, β, and materiality apply to the substantive testing procedure whereas α, combined risk, and materiality apply to the audit strategy. The simulation measures the actual α and combined risks which apply to an audit strategy, as well as the β risk of the substantive test.

At this point, it appears appropriate to describe in detail the audit strategies simulated. This is done by first describing the individual components of a strategy: internal control information, linkage rule, and substantive testing method; and then the different combinations of components that are used in the simulation.

4.3 Discussion of alternatives at each stage of an audit strategy

I Internal control information

Since the simulated accounting system is a closed system, that is all sources of error are determined by the system of internal controls, the simulation should be providing an upper bound or limit on the value of this information in terms of its impact on the amount of substantive testing. The typical form of this information for audit use is error rate data for critical attributes (compliance deviations) of the internal control system. Hence the simulated auditor is assumed to want to

[5]This relationship was first described on pp. 34-35 and footnote 8 of chapter two.

obtain this kind of information via the simulation of audit strategies.[6]

Since the basic goal of the simulation is to test the statistical validity of the internal control hypothesis (that is all tests are based on statistical sampling and conclusions are assumed to be consistent with statistical test results), a situation where all sources of nonsampling error are eliminated must be considered in the simulation. Therefore, in some simulations it must be assumed that when the simulated auditor obtains his compliance test results, he is able to accurately interpret it in light of the actual relationships of the simulated accounting system. This is reflected in most of the sampling strategies simulated.

The following represent the different amounts of internal control information used in the simulation of sampling strategies: (A) perfect information about the internal control system, (B) objective internal control information via statistical sampling for attributes using random sampling without replacement, (C) objective internal control information via statistical sampling for attributes using dollar-unit sampling, and (D) subjective information about the internal controls, and (E) no information. These will now be discussed further.

I(A). <u>Perfect information about the internal control system (omniscience)</u>. Discussion: The auditor is assumed to know exactly the error rates that apply to each critical attribute of the simulated accounting system. That is, for the low monetary error rate environment

---

[6]For examples of the kind of attribute data used by auditors please see footnotes 3 and 16 of chapter three and SAS No. 1 Sec 302B.24.

with reliability (probability of no monetary error occurring in an account) R = .99, the compliance error rate ϕ for each attribute is computed as follows:

$$(.99)^{1/5} = .99799 = 1 - \frac{\phi_1}{3} \text{ implies } \phi_1 = .006.^7$$

For the medium error rate system, R = .95

$$\text{implies } (.95)^{1/5} = .98979 = 1 - \frac{\phi_2}{3} \text{ implies } \phi_2 = .03.$$

For the high error rate system R = .91

$$\text{implies } (.91)^{1/5} = .98132 = 1 - \frac{\phi_3}{3} \text{ implies } \phi_3 = .056.$$

For the critical error rate system R = .90

$$\text{implies } (.9)^{1/5} = .97915 = 1 - \frac{\phi_4}{3} \text{ implies } \phi_4 = .063.$$

For the very high error rate system R = .85

$$\text{implies } (.85)^{1/5} = .9680 = 1 - \frac{\phi_5}{3} \text{ implies } \phi_5 = .096.$$

The symbol ϕ_i represents the compliance error rate that exists for each of the five attributes (represented by the subscript i) for the five environments. Note that all individual attribute rates are less than .10; this means that the Poisson distribution should provide a good approximation to the statistical measures of the sample results.[8]

---

[7]This computation follows from the fact that all component monetary error rates are equal (See assumption 1, p 121 of chapter three and pp. 122-32 of chapter three.) and hence that all the compliance error rates are equal (because of the same three to one ratio assumed for all critical compliance deviations). So using the formula for reliability $R = (1-\frac{\phi}{3})^5$, one can easily solve for the implied compliance deviation rate ϕ for a given R value.

[8]Pains have been taken to use compliance error rate ranges which appear to exist in practice, and which have implications consistent with those observed in practice. For example, see p. 40 and footnote 14 of

How the simulated auditor uses this actual error rate information is determined by the linkage rule used in a sampling strategy. Hence, further discussion of this is put off for later in the chapter.

I(B). Objective information about the internal control system through use of statistical sampling for compliance testing (i.e., sampling for attributes). Discussion: In deciding on how a compliance test sample size should be computed, it is necessary to consider several issues in making assumptions about the simulated auditor. First, it is assumed the auditor plans sample sizes for each of the five attributes assuming they all have an equal contribution in terms of the monetary error rate on the recorded book values. This will result in a conservative strategy (consistent in minimizing the risk of Type II error) as shown in appendix II. Since this is an exploratory study in this area (the evaluation of systems of internal controls) it appears best to use a conservative approach which is consistent with the conservatism present in other aspects of the audit (e.g., the conservatism inherent in the DUS philosophy and some of the linkage rules).

However, it should be noted that this conservatism in approach assumed for the simulated auditor is entirely eliminated in the simulation study because (1) the actual system structure is a series system, and (2) the actual component compliance and monetary error rates for each environment are equal. Much of the preceding discussion and

---

chapter two and chapter three footnote 16.

The Poisson approximation has been justified in footnote 11 of chapter four, and on pp. 36-40 A. D. Teitlebaum, D. A. Leslie, and R. J. Anderson, "An Analysis of Recent Commentary on appendix II thus applies to justifying the assumption of equal error rates in actual practice.

Given that the simulated auditor is aware the compliance error rates are equal, he still needs to test them in order to obtain sufficient assurance for reliance that the internal control procedures are being applied as prescribed. So the next issue facing the simulation of audit strategies is whether the compliance sample size should vary with the degree of reliance planned on the system. That is, generally speaking, the lower the compliance error rate the greater the degree of reliance the auditor can place on the system. However, as the error rate decreases and the system reliability increases, the larger is the sample size needed to obtain assurance that the error rate is at the specified level. This is clearly illustrated by the basic formula used by dollar-unit advocates: $n = \frac{B}{P}$ where n = sample size, B is the reliability factor associated with a given confidence level when using the Poisson approximation to the hypergeometric (e.g., for a confidence level of 95% this factor has a value of 3.0 for a discovery sample), and P is the critical error rate for determining the degree of reliance (also see footnote 20). Thus for 95% confidence, as the critical error rate changes from .05, to .03, to .01, the discovery sample size increases from $\frac{3.0}{.05} = 60$, to $\frac{3.0}{.03} = 100$, to $\frac{3.0}{.01} = 300$ (a fivefold increase over the given range). This fact results in somewhat of a paradox: if degree of reliance is related to compliance error rates

---

Dollar-Unit Sampling in Auditing," unpublished paper available on request from the authors, March 1975.

then increased reliance implies increased compliance testing. This fact is recognized by Don Roberts in his Statistical Auditing on p. 175 where he gives an example of sample size computations for compliance tests (customer orders) and various degrees of reliance thereon.

Thus the question is whether in simulating the auditor, compliance test sample sizes should be reduced to reflect possible less reliance for those situations where the internal control system is less than fully reliable.

On the other hand, one can argue that a given strategy should reflect a constant amount of information regardless of the state of internal controls and this information in the objective case appears to be best represented by a constant sample size across all internal control states. If the sample sizes were allowed to vary, one could justifiably argue the existence of an interaction effect between the level of internal control information and the linkage rule used to reflect that information. Therefore, to avoid such potential criticisms it seems most appropriate to use the same compliance test sample size for all the accounting environments. This is particularly true for testing the impact of statistical information on the quality of the internal control system. Also, in comparing two linkage rules for the same sample sizes, the relative effectiveness of each of the rules will be indicated for the same amount of sampling.

It is thus assumed that a fixed sample size is most appropriate for representing statistical information about the internal controls, and this same sample size will be used for all internal control states. This, of course, means that sequential sampling plans for attributes, which are not very common in auditing anyway (again, probably because the value of the information is so nebulous), will not be included in the simulation of audit startegies.

So the remaining issue with regard to the sampling plan for compliance testing is to decide on the sample size for the simulation of compliance tests. Again since there are five accounting populations, it is necessary for the compliance test sample size to be large enough to provide a fair degree of assurance of distinguishing among the associated compliance error rates. In particular the sample size should be large enough to provide a reasonable assurance of identifying the environment with $R = .99$.

It should be noted at this point that the compliance test sample size can be reduced considerably if it can be assumed all five attributes of the internal control system are obtainable from one document. Then sample sizes could be computed on the basis of the set of attributes as representing a particular aggregated attribute. Although this assumption would be feasible for testing compliance with procedures for classes of transactions, it is less realistic to assume this for an accounts receivable file with critical internal control procedures providing attributes from different supporting transaction files as discussed in appendix II. Thus it seems the most general approach to take is to assume each attribute (compliance error) must be sampled and estimated separately.

This assumption can result in a very large amount of total sampling for attributes and one may wonder a priori whether such extensive testing could ever be justified by any reduction in substantive tests.

However, it must be remembered that, typically, compliance tests can affect two levels of substantive tests--both at the transaction level and at the account balance level. Thus, for example, attribute test results associated with sales invoices can affect not only the extent of substantive tests for sales but for the ending accounts receivable balance as well.[9] Also, most samples are used for testing several accounting controls. (See the sales invoice example interpretation in chapter three.) In addition, the substantive test results may in turn affect other substantive and compliance tests. For example, discovery of a monetary error in the confirmation of accounts receivable should result in the "investigation of the cause of error with special emphasis on the control breakdown that permitted the error." "It is important to reevaluate the system of internal control" and this includes all the transactions relating to the system.[10]

Thus, because of the many accounting interrelationships, a given compliance test result can impact on several subsequent audit procedures. However, the same basic linkage rules are used in all these situations. Hence, by restricting the analysis to the impact on one substantive test only, this study provides evidence on the importance of the internal control information in general (that is, in reducing any substantive test, because it has been shown in appendix I and

---

[9]A compliance test, which typically applies to transactions, can affect the amount of substantive testing for transactions (i.e., the degree to which the testing is dual purpose) as well as the amount of substantive testing for the related account balances as discussed in footnote 3 of chapter three.

[10]Arens and Loebbecke, p. 196 and p. 334.

chapter three that the simulation pertains to either level of aggregation, transactions or account balances).

Since each of the attributes used in the simulation is independently distributed, it is possible to simulate an independent statistical sample by simply reading across the five fields associated with the five attributes and recording whether a zero or one value shows up for a particular attribute. The sampling for attributes is done simultaneously because this still ends up with a separate statistical estimate for the error rate associated with each attribute since the zero or one values for each attribute are independently generated. Because the error rates are equal, the same sample size is used to test each attribute as well as each accounting environment. What this sample size is, is discussed next.

Since none of the compliance error rates is over .10 and the sample sizes are at least moderately large (i.e., over 50), the Poisson distribution is a good approximation (slightly conservative) to the bionomial distribution.[11] Now, in planning for the fixed compliance test sample size to be used in the simulation, it is necessary to use a sample sufficiently large to be able to distinguish from the lowest error condition that arises in the simulated environments. This really means planning the sample size so that it is large enough to supply sufficient evidence for reliance when the highest degree of

[11]According to Winkler and Hays p. 232, if $n/p = \frac{50}{.1} > 500$ the Poisson approximation provides good accuracy to at least two decimal places. See Robert L. Winkler and William L. Hays, Statistics: Probability, Inference, and Decision, (New York: Holt, Rinehart and Winston, 1975). Also see p 155 of chapter five and footnote 8 of chapter five.

reliance is planned. This occurs, of course, for the population with the lowest error rate. This population is the one where R = .99 and so the associated compliance error rates are $\phi_1 = \phi_2 = \phi_3 = \phi_4 = \phi_5 =$ .066 (See p. 155 of chapter four). It is assumed that the auditor wants an overall confidence level for his set of compliance upper error limits of .95.[12]

The use of a given confidence level for a set of statistical tests (i.e., for the system monetary error estimate obtained from the set of five internal control tests) introduces statistical problems associated with estimating series system reliability from sample component data. As indicated earlier, several methods exist for estimating the lower bound on series system reliability. The simplest and most general method, assuming that the tests on the different subsystems give independent outcomes, is to use the fact that the probability of a joint event (the confidence level associated with the set of five compliance error tests) is the product of the individual probabilities (the confidence levels associated with each individual compliance error test). Thus a lower bound on series system reliability with five components and confidence level of $f_1 \cdot f_2 \cdot f_3 \cdot f_4 \cdot f_5$ is $\underline{R}_{f1} \cdot \underline{R}_{f2} \cdot \underline{R}_{f3} \cdot \underline{R}_{f4} \cdot \underline{R}_{f5}$ where $\underline{R}_{fi}$ is the lower bound on the i'th component reliability with confidence

[12]This appears to be a very common level of confidence for compliance tests actually used in practice. Examples of sources that have proposed this confidence level include SAS No. 1, Sec. 320B.24; Don Roberts in Statistical Auditing, p. 56, p. 62, p. 64, and p. 176; the Ernst and Ernst manual p. 9; and Arens and Loebbecke state on p. 289 of their text: "There is a general consensus in the profession that tests of transactions should range from 90 to 95 percent."

level fi.[13] This method for computing the lower bound on system reliability is henceforth referred to as the crude bound.

It turns out that this crude method for computing lower bounds on series system is quite general in the sense that it can be applied in the cases where the number of tests per subsystem are different, and the subsystem confidence limits can be based upon any kind of sampling, e.g., binomial sampling with sample size fixed in advance or number of failures fixed in advance. In addition it appears to fulfill all the requirements necessary for implementing an audit strategy. For example, the method is useful for planning the confidence level to use for each individual compliance test given the auditor would like a certain level of confidence for the lower bound on overall system reliability, the method can be used to readily implement all of the linkage rules used in the study, and the method assures keeping the risk of unwarranted reliance to at least the planned nominal level.

Unfortunately, as is shown in chapter five, the crude method proves to be too conservative for the case of the five internal controls used in the simulated accounting environments. This is true even for the highly reliable environment where such conservatism is minimized.[14]

---

[13]See Lloyd and Lipow, pp. 224-226. Again, lower bound on system reliability $\underline{R}_{fi}$ = (1-system upper error limit for i) at the confidence level fi used for the attribute test. See pp. 127-9 of chapter three.

[14]When a system is sufficiently reliable, the series assumption can provide a good approximation no matter how the system is actually organized (See Barlow and Proschan, p. 35). This provides yet another justification for making a series system structure assumption in auditing. However, the crude bound proves to be too crude for the reliability levels used in the simulation.

Hence a more accurate approximately optimum bound developed by Nancy Mann is used in the simulation of sampling strategies--it is fully described in appendix IV. The alternative Mann method has been chosen because it is relatively simple to apply and compares very well with other more accurate but more complex methods (but systems of only three or less components were used in the earlier simulation studies).[15]

The general problem with the more exact approaches is that they may overstate system reliability slightly for the two and three component systems studied in prior research. No results have been published for five component systems so it is not clear whether such overstatement gets worse. In fact none of the earlier studies compared actual system reliability to computed bounds, instead certain sample results were assumed and then various model estimates of lower bound system reliabilities were compared.[16] One advantage of the crude bound is that at least it always guarantees a bound at the stated confidence or higher.

Another possible disadvantage with the more exact approaches is that there is no longer a straightforward way of planning for the individual attribute test sample sizes as is the case for the crude

---

[15]See references in footnote 43 in chapter two.

[16]In these studies it has been generally assumed that a bound computed using the Buehler method is optimum. However these bounds can only be computed for up to three components. See footnote 55 of chapter two and Robert J. Buehler, "Confidence Limits for the Product of Two Binomial Parameters," Journal of the American Statistical Association, (December, 1957), pp. 482-493.

approach. The use of the Mann method, however, still allows the operationalization of all the linkage rules used in the study, and these are described later in this chapter.

It should be noted that the significance of the method used to compute the lower bound on series system reliability is reduced somewhat by the fact that strategies having perfect information about the internal controls are also simulated. The point of using sample information is to see how much of a reduction of an impact there is because of less perfect information.

Even though the Mann method for computing the lower bound on series system reliability is the one used in the simulation, the simple crude method is used as the basis for discussing auditing issues in this chapter. This is becuase it provides a simpler conceptual framework for discussing these issues. However, the Mann method follows a parallel logic and differs primarily in the mathematical distributions used (i.e., the Mann method works with the normal distribution as opposed to the Poisson approximation to the hypergeometric for the individual compliance error rates) although this results in much more precise estimates. Thus much of the following discussion, which is about the individual compliance test results more familiar to auditing, does not pertain to the same degree when using the Mann method. (For example, see chapter five for the actual risks of unwarranted reliance.)

Using the cruder and more general approach for computing the lower bound on system reliability and assuming a 95% confidence level for the set of statistical tests, requires that the test of each component be set at the 99% confidence level (i.e., $(.99)^5 = .95$). Hence one is really talking about very large sample sizes to detect an occurence rate of .006.[17] Several approaches are possible. Before these are listed it should be mentioned that most audit sampling plans and tables for attributes are designed for controlling the risk Type II error only, and even those advocating control of Type I error concede this is not practicable when the error rate gets below .01.[18]

Using this control of Type II error criterion as the basis for selecting the fixed sample size for compliance testing, the following approaches appear feasible.

Approach 1: Compute a discovery sample size for attributes for the lowest error rate $\phi = .006$ at 99% confidence.[19] The Poisson probability theory gives the sample size as n =

$$\frac{\text{Reliability factor for 99\% confidence and no errors}}{\text{error rate}} = \frac{4.61}{.006} = 769.^{20}$$

---

[17]See p. 155 to see how this error rate is obtained from the environmental relationships.

[18]References indicating only Type II errors are controlled for in attribute sampling include SAS No. 1 Sec. 320B.24. Most audit texts do not even mention α risk for compliance tests, e.g., Arens and Loebbecke or Robertson. Neither do the Ernst and Ernst, Clarkeson, Gordon, & Co., or Haskins & Sells manuals. Perhaps the reason for this is that, as Don Roberts points out, most auditing attribute tables are not designed for the control of α risk, p. 150 of <u>Statistical Auditing</u>. And this in turn is probably due to the fact that the value of additional internal control testing is questionable (e.g., see Clarkeson, Gordon, & Co. manual p. 120).

[19]Again to obtain an overall 95% confidence level for the lower bound on system reliability as discussed on p. 162 of chapter four.

[20]What is called here the reliability factor for the Poisson distribution for confidence level of $1-\beta$ and K errors in the sample is the mean ${}_\beta P_k$ of the Poisson random variable X such that the probability

This is a very large compliance test sample size; however, it is necessary for detecting such a low error rate with such a high level of assurance. Note under this approach if the error rate is exactly .006 the given sample size will have at least 1 error 99% of the time. (The risk of concluding the error rate is less than .006 when it is not is 1%, i.e., 1% is the risk associated with making a Type II error for the error rate).

Approach 2: Recognize that it is only necessary to discriminate the error rate range that allows maximum reliance on the internal controls. As is discussed later in this chapter, this occurs when the reliability of the system is estimated to be greater than .95 (i.e., $R \geq .95$).. This implies that it is only necessary to be able to detect a compliance error rate of $(1-\frac{\phi}{3})^5 = .95$, implies $\phi = .030$.

Hence a discovery sample size for detecting this error rate with 99% confidence is $n = \frac{4.61}{.03} = 154$. However, if the error rate in the population is actually .006 then this discovery sample size has an α risk (using the Poisson approximation) of Prob [more than 0 errors in 154 sample items given that the error rate is .006 or less] = $1 - e^{-154(.006)} \times \frac{[(154)(.006)]^0}{0!} = 1 - e^{-154(.006)} = 1 - .39693 = .60307 = \alpha$. This is the risk that the compliance test will indicate the error rate is greater than .03 at 99% confidence when the actual error rate is only .006. The number .03 represents the maximum error

---

that $X \leq K$ is β (i.e., Prob $(X \leq K) = \beta$). So the reliability factor for 99% confidence and no errors = ${}_{.01}P_0$ = 4.61 is the Poisson mean such that Prob $(X \leq 0) = .01$. This reliability factor ${}_{\beta}P_K$ is well approximated by the product n·(error rate) when n = the sample size is not small and the error rate is small. Thus the formula for the sample rate that would allow the greatest degree of reliance on internal controls. This is, if the sample indicates the compliance error rate is less than .03 with a 99% confidence level, then maximum reliance on internal controls would be placed; otherwise, reliance would be reduced (but not necessarily eliminated).

Approach 3: A sample size intermediate to the two extremes above is computed by using an approach followed by dollar-unit sampling advocates which is also based on attributes sampling theory using the Poisson approximation to the binomial distribution. Basically, their approach to sample size planning is to start with a prespecified upper error limit (which for purposes of discriminating the lowest range of error rates for maximum reliance must, therefore, be .03) and to make adjustments for the anticipated error rate (which for the most reliable population is .006) and expected "precision gap widening factors" (increasing precision when errors are discovered, which apparently is usually below 50% of the expected error rate).[21] To compute the interval to be used in the sample size calculation, then, these adjustments are subtracted from the prespecified upper error limit: Thus, if for convenience precision gap widening factors are set = .003, the interval to be used in the sample size calculation is .03 - .006 - .003 = 021. This .021 value is then used to compute the sample size: $n = \frac{4.61}{.021} = 220$. Note this is no longer a strict discovery sample because the acceptable

---

size comes from the relationship ${}_{\beta}P_K = n\cdot(\text{error rate})$ which implies $n = \frac{{}_{\beta}P_K}{(\text{error rate})}$. See pp. 42-45 of the Clarkeson, Gordon, & Co. manual.

[21]See pp. 18-21 and p. 32 of the Teitlebaum, Leslie, and Anderson paper.

upper error limit is still up to .030 (the discovery sample size is 154 as computed earlier).

Now consider the α risk associated with this plan when the actual error rate is .006. Again, using the Poisson approximation note that, if no errors are discovered in the sample, the upper error limit at 99% confidence is $4.6 \cdot x \frac{1}{220} = .02095$; and for one discovered error it is $6.64 \, x \frac{1}{220} = .03018$. This could still be accepted if the conservative approximation .03 were not used. So let n = 225, then the upper error limit is $6.64 \, x \frac{1}{225} = .02951$. With two errors the upper error limit is $8.41 \, x \frac{1}{225} = .03738$; therefore, the proposition that the error rate is less than .03 would be rejected with 99% confidence if two or more errors appeared in a sample of 225.

The α risk associated with this sample if the error rate is actually .006 (which it would be for the most reliable accounting environment) is therefore: Prob {more than 1 errors in 225 samples given that the error rate is .006 or less} $= \alpha = 1 - e^{-(225)(.006)} \sum_{j=0}^{1} \frac{[(225)(.006)]^j}{j!}$

$= 1 - e^{-1.35}(\frac{(1.35)^0}{0!} + \frac{(1.35)^1}{1!}) = 1 - .2592(1+1.35) = 1 - .6092 = .39079 = \alpha$.

It is evident that to reduce the α risk significantly for the low error rate conditions would require a substantial increase in the compliance test sample size. On the other hand, the advantages of the additional sampling required are so very nebulous at the present time that most auditors ignore the α risk altogether in attribute sample size computations.[22] It must be noted that the penalty for rejecting a .006 population and assuming it is really a .03 error rate or higher condition is simply in reducing reliance on internal controls. At worst such a rejection might cause more extensive substantive tests. So the trade off is really in terms of expanding compliance testing. Since substantive tests give much more direct evidence of the accuracy of financial statements, auditors have naturally tended to disregard α risks for compliance tests.[22]

This still leaves the problem of the sample size to use for compliance tests. Compared to examples given in the audit literature, the sample sizes given above are large. Only approach (2) results in a sample size comparable to published examples for compliance tests.[23] However, it is also true that there has been little recognition given in the audit literature for the problems associated with evaluating the results of a set of audit procedures (e.g., a set of compliance tests as is the case here). The use of the reliability model gives explicit recognition to this problem by adjusting the confidence level for each separate test (.99) so that a given confidence level (.95) will apply to the entire set of tests. It is primarily due to this unusually high confidence level (.99) for the individual compliance tests that such

---

[22]See references in footnote 18.

[23]For example, see Arens and Loebbecke, p. 302, where attribute sample sizes range from 50 to 150; or Don Roberts, p. 57 or p. 175, where sample sizes are from 50 to 235. On the other hand, Robertson's attribute sample sizes are much larger, p. 268, where the sample sizes are from 220 to 900.

large sample sizes result. If the more typical 95% confidence level would be used for each individual compliance test, the resultant sample size would be $n = \frac{3.0}{.006} = 500$ for approach 1, $n = \frac{3.0}{.03} = 100$ for approach 2, and $n = \frac{3.0}{.021} = 143$ for approach 3.[24]

An argument against using a very large compliance test sample size is that a strategy having perfect knowledge of the error rates is also simulated and thus using very large compliance tests sample sizes appears redundant.

On the other hand, a certain minimum sample size is necessary for use with the Felix and Grimlund model which requires a minimum of three observations of errors to be reliable. In the low error rate environment this provides an argument for using a compliance test sample size of about at least 150 to 200 (so that there is a reasonable chance of finding three or more errors in all situations).

Given the various issues involved it appears the sample size for compliance tests is somewhat arbitrary, but it is the feeling of the researcher that a reasonable fixed sample size for compliance tests is 150. The reasons for choosing this sample size are the following.

1. This represents an intermediate sample size for many audit tests and so it certainly is reasonable for most situations. In fact the much larger sample size of 300 is the discovery sample size when the auditor wants 95% assurance that the error rate is not greater than .01.

---

[24] Robertson obtains much larger sample sizes for attributes by using a confidence level of .99, see his Exhibit 7-6, p. 268.

2. This sample size should make it possible to implement the Felix and Grimlund model by not violating the assumptions of their model too severely (i.e., that there be a reasonable chance that at least three errors be present in using the model for preposterior analysis).

The big argument against using this sample size is, of course, that the α risks may be too high for linkage rules that depend on error rate estimates. For example, the α risk associated with this sample size when the actual error rate is .006 and the critical error rate for determining the degree of internal control reliance is .03, is about $\alpha = .60307$ as computed on p. 167 of this chapter. However, for such low error rates it takes very large sample sizes to control for small α and β risks simultaneously--sample sizes much larger than auditors ever contemplate. The reason for this is the uncertainty of the benefits associated with this additional information. Note, however, that as the critical error rate gets higher for lesser degrees of internal control reliance, the α risk will drop accordingly. Most importantly, since the impact of perfect information about internal controls is also ascertained, the impact of imperfections of the statistical sampling information is capable of some assessment.

For these reasons, the researcher argues that a fixed compliance test sample of 150 appears reasonable.

The objective information about the internal control system therefore consists of a compliance test sample sizes of 150 items which are used to obtain statistical information about the five attributes representing the internal control system. As discussed on pp. 156-9 of this chapter, this same sample size is used for each of the five attributes in each of the five accounting environments. However, these 150 sized samples are taken two different ways depending on the linkage rule used.

I(B). Objective internal control information via statistical sampling of records. Discussion: Under this approach the samples of 150 for attributes is obtained by systematic sampling of the accounts or records. In order to assure that the sample is a true random one without replacement, the records will be randomly reordered at the outset. This approach provides statistical information on the compliance error rate per record.[25]

I(C). Objective internal control information via statistical sampling of dollars. Discussion: Under this approach the samples of 150 for attributes are obtained by systematic sampling of the dollars associated with the records. This provides statistical information on the compliance error rate per dollar.[26]

---

[25]Systematic sampling is described on pp. 20-21 of Don Roberts' Statistical Auditing. Under the conditions of the simulation, a systematic sample is essentially equivalent to unrestricted random sampling. Many auditors appear to use systematic sampling as a matter of course in their audits, particularly the firm of Haskins and Sells.

[26]That is, attribute data is obtained by using proportional sampling (see pp. 95-97 of chapter two) or DUS. Again, since the accounts are first randomly scrambled in the sampling process, a systematic sample results in a virtually assured unrestricted random sample (see footnote 26).

It should be noted that this difference in sampling for attributes (i.e, one is a random sample of dollars associated with the records) rerults in a rather subtle difference in interpreting the resultant error rates: alternative I(B) results in a lower bound on the proportion of

I(D). Subjective information about the internal controls. In alternatives I(B) and I(C) of level of information about internal controls, it is assumed no subjective errors enter into the auditor's judgment. That is, auditors' actions and interpretations are consistent with the rules governing the underlying accounting system. The compliance test sample information allows the auditor to make an assessment of the reliability of the system and this, in turn, allows him to decide how much, if any, to reduce substantive testing. In computing this reliability, he must have some knowledge of the underlying account account processing system and in alternatives I(B) and I(C) it is assumed the auditor has sufficient expertise to translate the sample results in terms of actual system reliability. However, in the real world the auditor is likely to make judgmental errors in interpreting the sample results in terms of the impact on the reliability of the internal control system; and, hence, this will have an affect on the extent of subsequent audit procedures to be performed. Empirical evidence is available that auditors experience a high variance in assessing the impact of internal control information.[27]

In order to help assess the significance of introducing these judgmental errors, it is proposed that errors be introduced in the simulation of the auditor's estimate of system reliability. This can be done by picking a randomly generated number, $\tilde{\hat{R}}$, from a normal

---

of records not having any monetary errors; while I(C) results in a lower bound on the proportion of dollars not having any monetary errors.

[27]The empirical work by Joyce, Weber and Mock reviewed in chapter two indicates such judgmental errors occur frequently in practice.

distribution centered at the estimate, $\hat{R}$, of R obtained from alternative I(B) or I(C), whichever performs better, and substituting the randomly generated $\hat{\hat{R}}$ value. This normal distribution will have a mean of $\hat{R}$ and a standard deviation equal to 1/3 $\hat{R}$ and hence will vary depending on each $\hat{R}$ obtained every time a sampling procedure is simulated. The purpose of doing this is to obtain some assessment of the performance of an audit strategy when random judgmental errors are introduced to the process. Thereby some idea of the impact of factors which reduce such errors, such as greater auditor training and use of more sophisticated internal control models, can be obtained.

The use of a normal distribution to simulate the errors in judgment is based on arguments by Winkler and Hays.[28] According to the these two, statistical theory can be used to show that a random and independent error component from a complex process will tend to be normally distributed. It is assumed here that such errors characterize the errors in the estimation process.[29] The standard deviation may be somewhat large for this purpose but it is the intent of the study to get an upper bound on the impact of the reduction of such error.

This introduction of random judgmental errors is done for only one strategy—that which performs better than any of the others. Therefore, this subjective internal control evaluation strategy is the

---

[28]Robert L. Winkler and William L. Hays, Statistics: Probability, Inference, and Decision, (New York: Holt, Rinehart and Winston, 1975), pp. 251-252.

[29]As discussed in chapter one there is little behavioral data on which to build a modeling of judgmental error distributions. Hence purely mathematical arguments are used to justify the normality assumption.

last one simulated.

I(E). No information about the internal control reliability. Under this approach the audit sampling strategy consists of the substantive testing method only, with target values set at α = .05, β = .05 and materiality set equal to .05 of book value; i.e., no reliance on internal control is attempted and hence no error rate information is needed. The strategies using this approach provide the benchmark for comparing the effectiveness of having internal control information.

II Linkage rules

As discussed in chapter two, the reliability approach is used in modeling the internal controls of the auditing environments. This holds for modeling the audit sampling strategies as well as the accounting environments. Thus the linkage rules are all based on particular reliability levels which in turn are based on the compliance test results. The following linkage rules are used in the simulation:
(A) Elliott and Rogers' linkage, (B) SAS No. 1 linkage, (C) Clarkeson, Gordon, & Co. linkage, and (D) Felix-Grimlund's Bayesian linkage. These linkages are described in the following subsections.

Linkage II(A). The definition of reliability used for this alternative is the probability, R, that an account (record) is processed without a monetary error. This happens to represent the most conventional audit approach to interpreting the compliance test results when using reliability theory. For example, the most authoritative support for such an approach is stated as follows:

(a)...The committee believes that samples taken for this purpose should be evaluated in terms of the frequency and nature of deviations from any procedures the auditor considers essential in his preliminary evaluation of internal control, and that their influence on his final evaluation of internal control should be based on his judgment as to the effect of such deviations on the risk of material errors in the financial statements. [Emphasis added]

(b) The precision limits discussed in this paragraph for compliance tests relate only to deviations from pertinent procedures, which may or may not result in substantive errors in the accounting records (see paragraph 19) and

(c) Based on considerations of the general matters discussed in paragraphs 19 through 21 and of the specific factors mentioned in this paragraph, an auditor may decide for example that an upper precision limit of 10% for compliance tests would be reasonable; if substantial reliance is to be placed upon the procedures, he may decide, for example, that a limit of 5% or possible lower would be reasonable.[30]

The key points in this passage are the emphasis on frequency measures and on basing reliance on the upper error limit. As pointed out in chapter two, there are two general interpretations used in practice consistent with this view: the rigid methods and the flexible methods.[31]

Obviously, the more sophisticated conceptual approaches are the flexible methods because they attempt to reduce the inherent conservatism of the rigid approaches by letting the interpretation of a particilar compliance error rate depend on the particular accounting environment facing the auditor. This is more in the Bayesian spirit of

---

[30]AICPA, Auditing Standards, Sec. 320A.22 and 320B.22.

[31]See pp. 39-42 of chapter two. It should be noted that rigid rules may apply not only to a particular kind of attribute for all systems but for all attributes as well. For example, the rate interpretations given on p. 40 of chapter two would apparently apply to all critical compliance deviations in all systems.

SAS No. 1 as quoted on pp. 34-36 of chapter two, anyway, because, ideally, the auditor should be using his judgment to establish an accurate relationship between materiality as it affects tests of compliance, and materiality as it affects the financial statements or account balances; and this will depend on the particular situation at hand.

Since the simulated accounting environment establishes such a relationship, it appears most logical to devise an audit strategy that assumes an accurate assessment of the underlying relationships (i.e., assuming no judgmental errors are allowed to affect the strategy except in the case of alternative I(D) for internal control information described on p. 174 of this chapter). Thus the error rate translations are consistent with the simulated accounting environment. In particular this means the simulated auditor recognizes that a reliability of R = .9 implies an exactly material amount of total dollar error, and that he recognizes that each compliance error has a 1/3 probability of producing a dollar error in the records.

When using the reliability approach, the auditor must translate lower bounds on system reliability instead of upper bounds on compliance error rates. Mathematically the two translations are equivalent as pointed out on pp. 127-28 of chapter three. This can readily be seen for a subsystem with one attribute; a lower bound on subsystems reliability, $\underline{R}i$, (in the described system) for a given confidence level $= 1 -$ (upper error limit on i)/3 $= 1 - \frac{(UELi)}{3}$ for the same level of confidence on the upper error limit for attribute i. Similarly, a lower bound on series system reliability at a specified confidence level is comparable to the set of upper bounds of the associated independent compliance error rates associated with the system; e.g., lower bound on R = $(1 - \frac{UEL1}{3})(1 - \frac{UEL2}{3})(1 - \frac{UEL3}{3})(1 - \frac{UEL4}{3})(1 - \frac{UEL5}{3})$ at confidence level equal to the product of the confidence levels associated with UEL1, UEL2, UEL3, UEL4, and UEL5 where UELi represents the upper error limit for compliance error rate i.[32] (Division by 3 is necessary because the simulated accounting system will generate a dollar error for every 3 compliance deviations--the size of the dollar error is another matter.) Therefore, what are needed are strategies that can translate lower bounds on system reliabilities (whether calculated using the crude method or more accurate methods such as the Mann method described in appendix IV) to the degree of reliance on internal controls.

For alternative II(A) of the linkage rules a heuristic approach introduced by Elliott and Rogers is used, but refined for application on lower bounds on system reliability. A rule consistent with the underlying relationships of the simulated accounting environment is constructed. That is, for the system reliability R = .9 that results in an exactly material (.05 of book value) amount of error no reliance is planned, and as the lower bound estimate creeps up greater and greater reliance is planned.

Under the Elliott and Rogers' approach only a finite number of

---

[32]This formula results from using the crude method for estimating the lower bound on series system reliability discussed on p. 162 and footnote 13 above.

degrees of reliance are considered and hence all possible internal control states are categorized into a finite number of classifications. Others have supported this philosophy.[33]

> Since small variations in the degrees of reliance do not have much effect on the resulting extent of statistical tests, it is only necessary to select a few values from the range to express the possible degrees of reliance. For instance, the auditor might confine the possible degrees of reliance to 0, .3, .5, and .7 when the range is from 0 to .7. If the auditor does not want to use numbers, a qualitative scale may be substituted such as none, little, moderate, and high, to express the possible degree of reliance.[34]

Since the Elliott and Rogers' approach has become a classical work in audit research and predates the Robert's work, their approach is used in the simulation. Their approach involves a subjective evaluation of error rates and classification into excellent, good, fair, poor, or non-existent compliance with the system (i.e., five categories as opposed to Roberts' four). In the strategies where judgmental errors are eliminated and as a result of integration of the upper error limits on compliance deviations, it is possible to construct comparable rules relating to lower bounds on system reliability. The arbitrariness can be further reduced by using monetary error rate interpretations to qualitative scales made by others.[35]

As a result of considering these factors the following linkage rule

---

[33]These include Robertson; the Clarkeson, Gordon & Co. manual, pp. 120-121 and pp. 159-160; and the Haskins and Sells manual, frames 3-122 and 3-123.

[34]Don Roberts, Statistical Auditing, p. 132.

[35]For example, see Robertson's interpretations on upper error limit limits of compliance deviations, p. 367 and p. 369; Neter and Loebbecke monetary error rate interpretations, p. 127; and Don Roberts' interpretations on compliance error rates, p. 138.

is used. If the estimated 95% confidence level lower bound, $\hat{\underline{R}}.95$, on series system reliability is such that $.95<\hat{\underline{R}}.95\leq 1.00$, then $\beta = .5$; if $.93<\hat{\underline{R}}.95\leq .95$, then $\beta = .3$; if $.91<\hat{\underline{R}}.95\leq .93$, then $\beta = .15$; if $.90<\hat{\underline{R}}.95\leq .91$, then $\beta = .1$; and if $\hat{R}.95\leq .9$, then $\beta = .05$; where β is the Type II risk level set for the substantive test.[36] The definition of $\underline{R}.95$ is the lower bound of system reliability at 95% confidence where reliability is the probability that a record is processed without a monetary error. Thus, using the crude method and letting UEL.99 (i) = the upper error limit for attribute i at 99% confidence as a result of taking the sample of 150, the following is a lower bound:

$$\hat{\underline{R}}.95 = \left(1 - \frac{\text{UEL}_{.99}(1)}{3}\right)\left(1 - \frac{\text{UEL}_{.99}(2)}{3}\right)\cdot\cdot\cdot\left(1- \frac{\text{UEL}_{.99}(5)}{3}\right)$$

(slightly more conservative because $(.99)^5 = .951>.95$)

This is true because the underlying accounting system is a series system where each compliance deviation has a 1/3 probability of resulting in a monetary error, where the dollar error generating process is such that when R = .9 there exists a material amount of net overstatement = .05 x book value.

Now, as shown in chapter five, the preceding formula is overly conservative (i.e., the lower bound can be significantly increased) for

---

[36] It should be noted that these interpretations have been adjusted to eliminate much of the conservatism of an Elliott and Rogers type linkage by allowing maximum reliance (i.e., $\beta = .5$) for half of the range (0-.05) of acceptable reliabilities. Of course, any conservatism can be completely eliminated by considering a two state world: complete reliance or no reliance with cutoff taking place at exactly .9. However, this would not be in the spirit of the gradualism expressed, for example, by the Elliott and Rogers' grades of internal controls or in the quote from Don Roberts given on p. 180 of chapter four. Thus the researcher feels it is necessary to reflect this gradualism by allowing maximum

computing the lower bound on system reliability in the actual simulation, and so the Mann method is used instead to compute the 95% confidence bound (see appendix IV). The above cruder reliability formula is used here mainly to illustrate the conceptual relationships between system reliability and compliance errors in computing bounds on the reliability.

Linkage rule II(B). So far the dissertation has worked with only one audit concept of reliability, that being the probability that a record is processed without a monetary error occurring. This is only one specific definition that follows from the most general definition of the term as applied to audit applications: the probability of correct processing. Note that the general definition allows many variations of meaning which can be used depending on their usefulness. In fact much more might be involved because the various linkage rules imply different definitions. Hence, it might be found that certain definitions give internal control information more value. This evaluation is one of the lesser goals of the proposed study for it might be the case that how the linkages are modeled can have an important bearing on the performance of an audit strategy.

Other ways of defining reliability for audit purposes follow (including the symbolic representations following in parentheses).

---

reliance over a disproportionate range. The two state—reliance, no reliance—linkage philosophy is fully expressed through linkage II(B) which is described next. The implications of the gradualism of a Type II(A) linkage are explored and discussed in chapter five.

1. The probability that a record is processed without a compliance error (RC)

2. The probability that a dollar in the file of records is processed without a monetary error occurring (RD)--the upper error limit on this is computed using alternative I(C)

3. The probability that a dollar is processed without a compliance error (RDC)--the upper error limit on this is computed using alternative I(C) sampling.

4. The probability that the accounting system does not result in a material amount of total dollar error (C)

The last three reliability concepts are considered in the remaining linkage rules, the first has been considered implicitly in the first linkage rule. Apparently prior researchers have not appreciated the extent to which these definitions can be used to implement certain linkage rules.

Unlike the first linkage rule, alternative II(B) allows for continuous degrees of reliance on internal controls. It is represented by the following formula from SAS No. 1 Sec. 320B paragraph 35 which has been previously introduced:[37]

$$\beta = \frac{(1-R)}{(1-C)} = 1-S$$

where β = the sampling risk associated with making a Type II error as a result of substantive testing

R = combined reliability level desired (1-R = combined risk level)

C = reliance assigned to internal accounting control and other relevant factors

S = reliability level for substantive tests (using the negative approach as pointed out by Don Roberts)[38]

---

[37]See p. 35 of chapter two.

[38]Note the term reliability used in <u>Auditing Standards</u> is synonymous with the statisticians' use of the term confidence level when the

To better understand the meaning of this formula, it is best to clarify the meaning of variables R and C in the auditing context. R is perhaps best defined by looking at its complement, the ultimate or combined risk. Essentially, the combined risk is the risk of not detecting material errors as a result of the application of the audit process. The simulation reduces this process to the application of an audit sampling strategy to a particular account where all the sampling is done on a statistical basis. (There is nothing in the formula to prevent the auditor from doing all his tests judgmentally, however, then the various risks cannot be controlled for objectively using statistical sampling theory.)

This combined risk is controlled for at .05 as stated earlier, β is determined from the formula (this is in effect the linkage rule) so all that needs to be specified is how C is to be computed in the simulation of audit strategies. C must obviously be related to the 150 item sample size compliance test results, but how does one account for the "other relevant factors" in the definition?

In developing the environments, a closed system is constructed so that all the sources of dollar errors are due to compliance deviations. Typically, the auditor would not bother testing compliance if for other reasons he felt he could not rely on the internal controls. These other factors include the possibility of management override of the system and the design of the system. However, the fact the simulated auditor is

---

negative approach is used. For example see Don Roberts pp. 40-48.

willing to test compliance, as indicated by the computation of the compliance test sample size of 150, presupposes that these other factors do allow reliance (in fact, by the way the simulation is designed, system design is perfect because without compliance errors there are no dollar errors generated).

The remaining "other relevant factor" is analytical review. However, analytical review is but another form of substantive testing and the major goal of the research is to study the impact of internal control on substantive testing. The substantive testing chosen here is the tests of details through statistical sampling. It is felt that by using only one form of substantive testing, a clearer picture of the impact of internal control information results. Also, while it is true auditors can avoid making an analytical review, they are required by audit standards to do a minimum amount of account balance testing (e.g., accounts receivable and inventories).[39] Hence tests of details can be argued to be more important forms of substantive tests than analytical review.

Before continuing on reliance on internal controls only, it should be pointed out that the same formula has been proposed for linking analytical review information to extent of substantive tests of details as is the reliance on internal controls. The equation that is used by some accounting firms and has even been given some official support in

[39]See AICPA, Auditing Standards, SAS No. 1 Sec 331.

Statistical Auditing by Don Roberts is the following:

$$\beta = \frac{(1-R)}{(1-C)(1-SP)}$$

where all variables are as defined earlier except C = reliance on internal controls only now, and SP = reliance on analytical review procedures.[40] Note that if there is no reliance on analytical review procedures, then only the reliance on internal controls can affect the level of β, and similarly for internal controls. The impact of a given amount of information about either (in terms of reliance) has the same impact on β. Now, if the simulation can show that for a given level of internal control information the equation on p. 183 is valid, then evidence will also be provided for the validity of the relationship for the same level of analytical review information. Thus the study also has relevance for the value of analytical review information.

Note that if no reliance is assigned to analytical review, SP = 0, the formula reduces to that of p183 with C redefined as above to pertain only to internal controls. This is the formula that is used in the simulation.

Thus a necessary step is to operationalize the variable C = reliance on internal controls so that it can be used in the simulation. This is a nontrivial task since no one has yet proposed an objective way of obtaining this value within the literal sense that SAS No. 1

[40]See p. 133 of Don Roberts, Statistical Auditing; or the Ernst and Ernst manual p. 70.

appears to intend. Of course, one can readily come up with a qualitative scale as was done in alternative II(A), but this approach really avoids the issue of what is meant by reliance and only establishes some rules for relating certain statistical results on error rates to β levels.

A review of SAS No. 1 indicates that the intended meaning of reliance on internal controls is really in terms of probabilities. For example, the quote on pp. 38 of chapter two indicates that degree of reliance is the complement of the risk that material errors will occur in the accounting process, i.e., a probability concept appears reasonable. Similarly, the quote on p. 177 of this chapter indicates compliance tests should be evaluated in terms of the risks of material errors that may arise. Again, interpreting this to mean "probability of material errors" appears to be reasonable. Additional evidence for this probability interpretation can be found in the following quote:

> "...the maximum degree of reliance on pertinent accounting internal controls ideally represents the auditor's assigned likelihood that the set of pertinent accounting controls would prevent or detect a material amount of monetary error."[41]

Likelihood is a conditional probability concept and again the probability interpretation of relaince appears feasible. In fact all these statements are not inconsistent with a Bayesian interpretation of sample results.

The Bayesian and perhaps likelihood interpretation of the attribute sample results provide an objective way of measuring an auditor's

---

[41]Roberts, p. 132.

reliance (and hence the "C" value) on internal controls.[42] This is because it turns out that a Bayesian interpretation of the compliance test results with a diffuse prior is consistent with the confidence level associated with the conventional statistical upper error limit. This is illustrated shortly.

Probably the best exposition of the Bayesian approach to linkage rules is given by Bailey and Jensen.[43] In essence their general Bayesian revision model is as follows:

$$(1)\quad P(F/Yi) = \frac{P(Yi/F) \cdot P(F)}{P(Yi/C) \cdot P(C) + P(Yi/NC) \cdot P(NC)}$$

where P(F/Yi) is the conditional probability of a fairly presented, F, account balance; Yi is the signal from the compliance test; C is the state that the accounting system is substantially in compliance and NC is the state that the accounting system is not substantially in compliance. They use a simplification wherein the state sets always consist of two elements, i.e., C and NC, and F (fair presentation) and NF (not fair or materially in error).

The generality of their model arises from the fact that they consider situations where $P(NF \cap C) \neq 0$ and $P(F \cap NC) \neq 0$. However, to deal with these situations they use the Bayesian revision given in equation (1) above. Notice that this differs from the usual Bayesian revision

---

42 For three different interpretations of probability see Morris H. DeGroot, Probability and Statistics, (Addison-Wesley Publishing Company Reading, Massachusetts, 1975), pp. 2-5.

43 Andrew D. Bailey Jr. and Daniel L. Jensen, "A Note on the Interface Between Compliance and Substantive Tests," Journal of Accounting Research, Fall 1977, p. 294.

in that a different state set is represented in the numerator (F, NF) than is represented in the denominator (C, NC). They justify this on grounds that equation (1) is consistent with the general form of the definition of conditional probability, $P(F/Yi) = \frac{P(F \cap Yi)}{P(Yi)}$ and that equation (1) results in more convenient probability estimates on the part of the auditor. This is because they feel the probabilities P(C/F), P(NC/F), P(C/NF), and P(NC/NF) are easier probabilities to assess than P(Yi/F) and P(Yi/NF); but they recognize that this is an empirical question open to research.

What implications does this have for the simulation? Well, a major advantage of the simulation methodology is the control available in defining the environment. More particularly, one can control for all sources of error in the account balance in such a way that the impact of internal control information is magnified and an upper bound on the value of this information is obtained. (The lower bound is zero because the auditor always has the option of not relying on internal controls.) This is accomplished by using a closed system, i.e., one in which all errors in the accounts arise as a result of compliance deviations. (In fact this appears to be implied by any auditor who tests for compliance because as discussed on p. 185, if the auditor felt there was a significant probability of error due to sources other than internal controls, he would not reduce the extent of substantive tests no matter how good the controls.) This effectively means that P(NC/F) and P(C/NF) would be zero or close to it whenever internal control reliance is being considered.[44] Since the only sources of errors are compliance deviations, and by definition, therefore, substantial compliance means fair presentation; $P(C/F) = 1$, $P(NC/NF) = 1$, and $P(NC/F) = P(C/NF) = 0$ in the simulated accounting environment. Thus assuming no judgmental errors, the general revision model reduces to $P(F/Yi) =$

$$\frac{P(Yi/F)\cdot P(F)}{P(Yi/C)\cdot P(C)+P(Yi/NC)\cdot P(NC)} = \frac{P(Yi/C)\cdot P(C)}{P(Yi/C)\cdot P(C)+P(Yi/NC)\cdot P(NC)} = P(C/Yi)$$

This is because it can be shown (see p. 296 of Bailey and Jensen): $P(Yi/F)=P(Yi/C)\cdot P(C/F)+P(Yi/NC)\cdot P(NC/F) = P(Yi/C)$ and $P(F)=P(F/C)\cdot P(C)+P(F/NC)\cdot P(NC) = P(C)$. Therefore, in a closed system $P(F/Yi)=P(C/Yi)$ and similarly $P(NF/Yi)=P(NC/Yi)$.[45] In other words, the probability of having a material error in the financial statements given the compliance tests results equals the probability of having compliance errors beyond the threshold level given the compliance test results. This assumes, of course, that the threshold compliance error level specified by the auditor actually results in a material amount of dollar errors being generated (i.e., no judgmental errors).

Thus in a closed system the probability of material financial errors reduces to assessing the probability that a certain set of compliance error rates or, equivalently, certain monetary error rates exist

---

[44]This is also reflected by the policy of Clarkeson, Gordon, & Co., for example, see p. 117 of their manual.

[45]This is consistent with the Clarkeson, Gordon, & Co. definition that "...every compliance deviation does not contain a monetary error... Every monetary error is; however, a compliance deviation." Ibid., p. 117.

in the accounting system. It so happens that the basic Bayesian approach for this is fairly straightforward in the continuous case if the right prior distribution (conjugate prior) is used. The only issue remaining to be settled is what form the prior distribution should take.

Since most of the audit strategies are not formally Bayesian it appears the most appropriate form of the prior distribution in all cases is that of a diffuse (informationless) prior. This puts the Bayesian (Felix and Grimlund) strategies on an equal footing with the non-Bayesian strategies which do not formally incorporate prior distributions, and thus a more relevant comparison is obtained. However, when one is referring to the evaluation of the compliance test results only, without yet considering the integration of the dollar amounts of errors, the diffuse prior distribution takes on a different importance as indicated by the following excerpt from Robert Winkler's book:

> When a diffuse prior distribution is used in a Bayesian analysis, the posterior distribtuion is virtually identical to the likelihood function, as pointed out in Section 4.10. Thus any inferences and/or decisions based on the posterior distribution will in reality depend almost solely on the sample information as summarized by the likelihood function. But a classical statistician bases inferences and decisions solely on the sample information. Under a diffuse prior state, then Bayesian and classical statistical procedures are based on the same set of information. If relevant prior information is available, the Bayesian's posterior distribution will reflect both this information and that of the sample and the Bayesian results are likely to be quite different from the corresponding classical results. In the special case of a diffuse prior distribution, however, classical and Bayesian results are quite similar, being based on essentially the same information."[46]

[46]Robert Winkler, Introduction to Bayesian Inference and Decision, (Holt, Rinehart, and Winston Inc., 1972) p. 388.

Winkler later goes on to illustrate for the normal distribution case that the significance level of the classical one tailed test is equal to the posterior probability associated with the null hypothesis using the Bayesian approach (Bayesians use their posterior distributions to make probability statements about the parameters of interest such as unknown population error rates, hence their interpretation of compliance test data appears very consistent with intended reliance on that data). It turns out that the same result holds for the binomial sampling distribution when a uniform diffuse prior is used over the range of error rates $0 \leq \phi \leq 1$.[47] This will now be illustrated by an example.

Assume a diffuse beta distribution (the conjugate prior to the binomial distribution) for the error rate p with parameters r' = 1 and n' = 2 (using Winkler's notation p. 201). Then it is easily shown that the posterior beta distribution has the parameters r" = 1 + r and n" = 2 + n where r = the number of compliance errors in a sample of size n.

Using the fractiles of the beta distribution tables in the back of Winkler's book and values of the Poisson process mean values used in the Haskins and Sells manual, one obtains the following figure of results assuming a discovery sample size of n = 100 and no errors are found:

---

[47]See, for example, Teitlebaum, pp. 11-12.

Figure 5

Bayesian with diffuse prior		Poisson approximation to the binomial	
r" = 1; n" = 102		n = 100 with no errors	
f	Pf	confidence level	upper error limit
.5	.00686	.5	.007
.75	.01372	.75	.014
.9	.02280	.9	.023
.95	.02966	.95	.03
.99	.04559	.99	.046

where Pf is a compliance error rate value such that $P(\tilde{p} \leq Pf/1, 102) = f$ and $\tilde{p}$ is the unknown compliance error rate; i.e., f is the probability that the unknown error rate $\tilde{p}$ is less than p_f which the auditor can define to be the threshold rate. Similar comparability is obtained when errors are found in the sample. Note that the Poisson approximation is always a little conservative (i.e., larger upper error limit).

The implication of this example is that the confidence level associated with attribute sampling would be interpreted by a diffuse prior Bayesian as the probability that the error rate is less than the upper error limit.[48] This means that the confidence level associated with a statistical test is a reasonable basis for determining reliance on

---

[48]Significance level is defined by Winkler p. 422 to be "the chance of obtaining a sample result as unusual as or more unusual than the one actually observed given that the null hypothesis is true." Since the null hypothesis in compliance testing is that there is a material error rate, the probability associated with this state of the world is the significance level = β; and the probability associated with the alternative hypothesis is $1-\beta$ = confidence level. Thus in the example given with the particular sample result of no errors in 100 items, the Bayesian auditor can make the following equivalent statements: he feels there is a 50% probability the error rate is less than .007, a 75% probability it is less than .014, a 90% probability it is less than .023, and so on. A similar statement can be made for any other upper

internal controls (i.e., the C value of the formula on p.186 of this chapter). Most importantly, it provides a completely objective basis for computing the number C that can be used in the simulation of sampling strategies.

It must be noted, though, that not every statistician would agree with this treatment of the confidence level. There are at least three schools of statistical thought which have different philosophies concerning interpretations of sample results.[49] It is not the intention of the researcher to get involved in the theoretical controversies of the Bayesians versus the non-Bayesians. Let it be said only that there exists considerable support for interpreting the confidence level of the compliance tests as given here, which is completely objective and consistent with the meaning given to reliance on internal controls by the auditing literature. Since this appears to be the best objective way of operationalizing the SAS No. 1 linkage rule literally, it is proposed for the simulation. Note that this will allow a valid comparison between the performances of the different linkage rules because the same compliance test sample is made available to all strategies and the simulated auditors are given the insight that R = .9 is the materiality threshold limit.

---

error limit including the exactly material error rate. This is the basis on which "C" is calculated.

[49]Winkler, p. 389, distinguishes between the sampling distribution interpretation, the likelihood interpretation, and the Bayesian interpretation.

The actual calculation of the reliance J given the compliance test results and using the crude method formula for computing the significance level is as follows. Noting that the exactly material system monetary error rate is .10, the associated reliability is R = .9, and the associated compliance error rate for each attribute i is ϕi = .063;[50] the confidence level Ci for K errors found in a compliance test sample of 150 represents the reliance (as argued in footnote 48 and p. 193) assigned to control i. That is,

$$Ci = 1-e^{-(150)(.063)} \sum_{j=0}^{K} \frac{[(150)(.063)]^j}{j!}$$

= prob (that ϕ_i is less than .063 given that K errors have been found in a sample of 150 items)

= prob (more than K errors in 150 samples given that the error rate is equal to .063). The second probability definition is the confidence level interpretation while the first definition is the associated equivalent Bayesian interpretation with a diffuse prior.

Since Ci is the probability that the error rate for attribute i is less than materiality, it also represents the reliance that can be placed on that control procedure. Similarly, the probability that none of the subsystems breaches materiality and, hence, the probability that the entire internal control system does not result in a material error is the product of the individual probabilities of immaterial compliance

---

[50]See p.155 for the calculations.

error rates, i.e., C = C1·C2·C3·C4·C5. Thus C is the assured conservative reliance that can be placed on the entire internal control system.[51] This is the "C" value which can be used directly in linkage rule II(B) on p. 186 of this chapter.

Unfortunately, as is shown in chapter five, the crude method formula proves to be too conservative for use in either linkage rule II(A) or II(B), and thus the C value used in the simulation is computed from the Mann method formula as described in appendix IV.

Finally, it should be noted that because of conventions that have developed in auditing, using the II(B) linkage limits the maximum β value allowed to $\beta = .5$.[52] Since the goals of the simulated auditor include control of combined risk at the .05 level, the minimum β (when there is no reliance on internal controls) is set at $\beta = .05$.
Thus with the above constraints, linkage rule II(B) consists of the following: $\beta = \frac{1-R}{1-C}$ where 1-R = combined risk level desired = .05, so that the formula in the simulation is $\beta = \frac{.05}{1-C}$ and $.05 \leq \beta \leq .5$.[53]

One might wonder given the Winkler quote on p. 191 of this chapter why the Felix and Grimlund model is being considered if both Bayesian and non-Bayesian models can end up yielding the same results as illustrated above. The reason for this is that although the auditor

---

[51]Conservatism in the sense shown in appendix II.

[52]See pp. 74 and 75 of chapter two for more discussion on this topic.

[53]This is the formula from SAS No. 1 Sec. 320B. 35 as discussed on pp. 183.

is assumed to have a diffuse prior in regard to his compliance tests, this is not the case for the substantive tests when he already is presumed to have internal control information. Thus the difference arises in how the prior information about the substantive test is presented. The Felix-Grimlund model makes formal distributional assumptions about the prior (internal control) information, whereas the non-Bayesian models use other methods for telling the auditor how much substantive testing needs to be done as a result of internal control information. In addition, the Bayesian model then provides a consistent way of incorporating the substantive tests results in a final statistical decision on account balance accuracy. The non-Bayesian models, on the other hand, are used essentially to determine the extent of substantive tests and these substantive tests, in turn, determine the final statistical decisions. Thus both the Bayesian and non-Bayesian methods attempt to accomplish the same purpose of integrating the various sample results to reach a statistical decision on financial statement accuracy, they just do it differently. So it will be interesting to see how these two categories of strategies perform.

Finally, it should be noted that the Felix-Grimlund model does not require use of a Bayesian framework (i.e., formal integration of prior information) because the mathematics is such that with sufficient error observations a conclusion can be reached on the basis of the substantive tests information only. (In fact this is true of all Bayesian models because of the availability of the diffuse (informationless) prior.) Thus, because of the completely different mathematical models that it utilizes, the Felix-Grimlund model can be considered a new estimator for substantive tests on a par with stratified mean-per-unit and DUS estimators.

Linkage rule II(C): dollar-unit sampling of attributes. This alternative linkage rule is a specialized one that has been proposed for use with DUS when DUS is used for substantive testing.

There actually appear to be two variations of this linkage rule used in practice, both are conservative and one is much more conservative than the other. The more conservative one is used by the firm of Haskins and Sells. It is more conservative because it automatically reduces reliance on internal controls as soon as the compliance error rate per dollar equals the dollar error rate considered material. Another reason this method is not considered in the simulation is that it is unclear exactly how reliance is reduced. Apparently, firm policy in this regard is specified in a table which is revealed only to the staff of the firm.[54]

The linkage rule that is simulated is the one used by the firm of Clarkeson, Gordon, & Co. of Canada. They are much more explicit in what they do. The basic rule is to reduce reliance to zero as soon as upper error limit on compliance deviations is three times materiality (i.e., set β = .05), and to set the β level for substantive tests at .2 if the upper error limit on the compliance error rate per book value dollar is

[54]See frames 3-122 and 3-123 of the Haskins and Sells manual.

less than three times materiality.[55]

Now, this method also proves to be conservative for the simulated environments because it relies on the internal controls fewer times than it could. This may at first glance appear to be surprising since the simulated accounting environment is based on the relationship that every compliance deviation has a 1/3 chance of producing a monetary error and this is precisely the relationship assumed by Clarkeson, Gordon, & Co.[56] However, a rather subtle additional assumption made by the Clarkeson, Gordon & Co. method is that each monetary error results in a 100% overstatement error on the dollar whereas in the simulation the tainting can be anywhere from a 100% understatement to 100% overstatement. When all this is aggregated over the entire population, it will be found that in the simulated environment approximately 27-28% of all dollars will have a compliance deviation associated with it (i.e., at least one of the five attributes will have a one value) at the exactly material amount of unreliability (1-R = .1); whereas linkage rule II(C) begins to reject reliance when only about 15% of all dollars have compliance deviations (under the Haskins and Sells approach rejection of reliance would occur at the 5% mark).

---

[55]See the Clarkeson, Gordon, & Co. manual p. 160.

[56]In fact the decision to construct the accounting environment this way was very much influenced by the smoke/fire analogy used by advocates of this approach--see footnote 30 of chapter three and pp. 117-121 of the Clarkeson, Gordon, & Co. manual.

The Clarkeson, Gordon, & Co. rule is recognized by its advocates as being conservative but perhaps the 100% overstatement assumption makes it more conservative than intended. Anyway, an assessment of this conservatism is made by comparing the performance of this linkage rule between the omniscient (I(A)) and the objective (I(C)) internal control information alternatives.

Another source of conservatism of using this method is that reliance on internal controls only increases β risk to .2 while under linkage rules II(A) and II(B) the β risk is allowed to go as high as .5.[57] Again, it will be interesting to see how these various assumptions interact to affect the performance of a sampling strategy.

Linkage rule II(D). This linkage rule is the one implied by the Felix and Grimlund model.[58] Briefly, the auditor is assumed to assess a posterior beta density $f_{\beta}(\rho/k,n)$ on the basis of the compliance test results where ρ is the unreliability in terms of the monetary error rate (i.e., $\rho = 1 - R$). This beta distribution for unreliability is then combined with dollar amount error generation process information, which is assumed to be a normal process, to yield the distribution for the total dollar error in the population, Π_T:

---

[57]As first indicated in footnote 71 of chapter two, many DUS users do not normally let the β risk climb beyond .2. The reasons for this are vague but they appear to be based primarily on the fact that the quality of evidence on internal controls is much more judgmental in nature than evidence obtained from a direct substantive test and, therefore, the degree of assurance from the test should be "reasonably" high. See the Clarkeson, Gordon, & Co. manual, p. 159, and Teitlebaum, Appendix III, p. 39.

[58]William L. Felix, Jr. and Richard A. Grimlund, "A Sampling Model for Audit Tests of Composite Accounts," Journal of Accounting Research Spring 1977, pp. 23-41.

$f_{\beta N}(\Pi_T) = \int_0^1 f_\beta(\rho/k,n) f_N(\Pi_T/a\rho,\ 1/b\rho)\, d\rho$ where $a\rho$ and $b\rho$ are the mean and precision of a marginal distribution for the total error amount in a population with monetary error rate of ρ. The $f_{\beta N}$ distribution is then approximated by an extended beta distribution using the method of moments.[59]

To operationalize their approach, it is first necessary to define how the prior $f_{\beta N}$ is computed right after the completion of the compliance tests. Since by definition

$$(2)\quad \rho = (1 - \frac{\phi_1}{3})(1 - \frac{\phi_2}{3})(1 - \frac{\phi_3}{3})(1 - \frac{\phi_4}{3})(1 - \frac{\phi_5}{3})$$

where ϕ_i is the compliance error rate for attribute i, a "good" estimate of ρ based on the sample results is the estimate obtained from using the maximum likelihood estimate of each ϕ_i and then using equation (2).[60] Call this maximum likelihood estimate of the system monetary error rate or unreliability, $\tilde{\rho}$.

Now, a theoretical problem arises by the fact that after compliance testing but before substantive testing, the five attribute samples alone provide a composite estimate on the net effect of the internal control system. However, a system output is not tested directly until the substantive test phase of the audit. But in implementing the Felix-Grimlund model, one must be able to convert this information to an equivalent prior substantive test sample size. Neither Grimlund

[59] All of the equations are explained and summarized in Appendix V including some minor corrections from the original source.

[60] For a discussion of maximum likelihood estimates and their importance see Mann, Schafer, and Singpurwalla, pp. 81-85. The maximum likelihood estimate for $\phi_i = \frac{x_i}{n_i}$ where x_i is the number of

nor Felix addressed this problem of operationalizing their model. Instead it was always assumed that the auditor could obtain a valid prior.[61]

Fortunately, the better developed reliability theory literature has a partial solution to this problem. Researchers have estimated how much weight to attach to the prior information via a pseudo sample size capable of calculation when binomial data on components is available (e.g., the reliability approach as defined in chapter two, is followed).[62] This weighting is computed using the following formula:

$$\frac{1}{n^*} = \frac{.5\left[\sum_{j=1}^{k}(1/nj) + \frac{1}{n(1)}\right]}{n^{(1)} \sum_{j=1}^{k} (1/nj)}$$

where n(1) is the smallest compliance test sample size and nj is the compliance test sample size associated with each of the K = 5 attributes, j.[63] n* is thus the prior weight attached to the internal

---

failures out of a sample of ni for the component i process.

[61] In fact this problem is surprisingly similar to the unanswered question left by the non-Bayesian linkage rules: how does the auditor establish the prior relationship between compliance errors and the total amount of dollar error?

[62] The theory gets rather complex and only the final result is given here. See pp. 518-524 of Mann, Schafer, and Singpurwalla for a discussion and references to basic work in this area.

[63] Mann, Schafer, and Singpurwalla, p. 521.

control information as a result of compliance testing and it equals the equivalent sample size in terms of direct system (i.e., substantive) tests of this prior information.

The equivalent number of observed system errors is thus $n^* \cdot \tilde{p} = K$ in an equivalent substantive test sample size of n^*. This information can now be used to drive the Felix-Grimlund model.

However, one more detail needs to be settled and that is how to specify the mean and variance of the K dollar errors that are assumed to be equivalent to the results of a substantive test. Note that this problem arises for both the reliability and Grimlund approaches to modeling internal controls because even if dollar error data is collected for subsystem components this information must be reducible to an aggregate systems test. For example, assume that by using a Grimlund approach it is found that out of a sample of 100 for each of the control points, one monetary error is found. The maximum likelihood estimate of system monetary error rates is $\tilde{p} = .05 = 1 - .95 = 1 - (.99)^5$ and $n^* = \dfrac{5}{.5(\frac{6}{100})} = 5 \cdot \dfrac{100}{3} = 166$. Then the equivalent number of system monetary errors is $K = \tilde{p} \;.\; n^* = (.05)(166) = 8.3$ which is 3.3 more errors than actually found in the internal control evaluation stage. What does one then assume about the mean and variance of these dollar errors? As far as the researcher can determine this is an unanswered question in reliability theory and in Grimlund's work, and so it is necessary to make an assumption about the simulated auditor's prior in this regard.

This assumption is consistent with the reliability approach: assume the auditor is fairly informed about the error size generating process at the system level and that he can at least accurately assess the mean and variance of such a process.[64] With this assumption it becomes evident that the mean and variance to be assumed for the K monetary errors should be the actual mean and variance of the underlying simulated environmental process. This is the assumption that is used in the simulation.

With this assumption it appears that a reasonable solution has been found to the problem of comparability of prior information between the Bayesian and the non-Bayesian strategies. The non-Bayesian strategies are in effect told what level of unreliability, 1-R, results in an exactly material amount of dollar error. The Bayesian strategy, on the other hand, never asks for this kind of information. But it does make the implicit equivalent assumption that the auditor has sufficient expertise to develop an informative prior; otherwise a diffuse prior would always be used for substantive tests (i.e., no value to internal control information). Thus it appears that one reasonable way to make the Bayesian and non-Bayesian strategies more comparable is to give the

[64] This is certainly a function of data availability, again. This kind of information can be justified on the basis of prior auditor experience with the firm. The auditor's working papers over the years may provide sufficient data to model the error size process. Again, the greater the decomposition, the less likely sufficient data would be available for error size modeling. See footnote 3 of chapter three.

Felix-Grimlund Bayesian model information consistent with underlying process.[65]

Having made this assumption, the prior beta-normal distribution, $f_N(\Pi_T)$, of total dollar error after compliance testing but before substantive testing, can now be mathematically specified. First, the actual n* value used in the simulation is specified, which because each compliance test sample size is 150, as discussed earlier, is

$$\frac{1}{n^*} = \frac{.5[\frac{5}{150} + \frac{1}{150}]}{150(\frac{5}{150})} = \frac{\frac{3}{150}}{5} \quad \text{implies } n^* = \frac{5}{3/150} = \frac{5 \times 150}{3} = \underline{250}.$$

This then represents the prior amount of substantive test sample information as a result of internal control testing where it is assumed the compliance test priors are diffuse.[66]

Since, as discussed previously, $\tilde{\rho}$ can be obtained from the compliance test directly, all the data is available for specifying the beta component f_β of the prior $f_{\beta N}$ via the relationship $K = \tilde{\rho} \cdot n^*$ as discussed earlier.[67]

With this specification of the prior f_β, it is next necessary to specify the normal component, $f_N(\Pi_T/a\rho,\ 1/b\rho)$, of $f_{\beta N}$ after compliance

---

[65] It should be stressed that the Felix-Grimlund model is truly Bayesian only when prior information via the compliance tests is combined with the substantive test information. When only substantive test information is involved, the Felix-Grimlund model effectively becomes a new non-Bayesian substantive test method as discussed on p.198 of this chapter.

[66] See p191 and p. 197.

[67] The two parameters K and n* are sufficient for specifying the prior standard beta distribution f_β. See, for example, Winkler, pp. 149-150.

testing but before substantive testing. This is where the assumption that the auditor can accurately assess the actual mean and variance of the error size generation process comes into play. Thus the prior $a\rho = \chi m\tilde{\rho}$ is assumed where χ is the population size and m is the actual mean of the individual error size process; and $b\rho = \chi[(\frac{k}{k-2})(\frac{1+k}{k})\upsilon]$ where υ is the actual variance of the individual error size process.[68]

The $b\rho$ formula assumes that $k \geq 3$. This is very unlikely for $n^* = 250$ and $\tilde{\rho} \leq .01$, which is the case for the most reliable of the simulated accounting environments.[69] When this assumption is violated, the simulation program automatically increases the amount of substantive testing to satisfy the mathematical needs of the model.[70]

With these rules, the entire prior $f_{\beta N}$ is thus specified for all environments.

---

[68] These statistics are defined on a per dollar as opposed to a per record basis. This is discussed shortly.

[69] Remember, the most reliable environment is R = .99 which implies a monetary error rate or unreliability of $\rho = .01$. See p. 139 of chapter three for a comparison of the environments. The limitation is first recognized by Felix and Grimlund on their p. 30.

[70] One consequence of this mathematical limitation for low K values is the paradoxical result that more substantive testing is required for very reliable environments. Grimlund's apparent explanation for this limitation is that for either very high or very low reliability accounting systems, there may be no need to formally model the internal controls. See Grimlund, "A Framework for the Integration of Auditing Evidence," p. 103.

With the prior (after compliance testing but before substantive testing) $f_{\beta N}(\Pi_T)$ specified, it is possible via a preposterior analysis to compute a sample size n for controlling a probability of Type II error at the specified level of .05. According to Felix and Grimlund, this is accomplished by "determining the probability that the total error amount is outside some materially range" (in the case of the simulation this is the probability associated with having a total overstatement error greater than or equal to .05 of the book value).[71]

The procedure for picking the optimal n (in effect the Felix-Grimlund linkage rule) is outlined as follows:

Step 1. Using the prior $f_{\beta N}(\Pi_T)$ after compliance testing, compute via the extended beta approximation inverse procedure whether the probability of a material overstatement (.05 of book value) is greater than .05.[72] If it is, go to step 2; if it is not, set n = 0, i.e., there is no need to apply a substantive test and the accounting population is accepted.[73]

Step 2. Increment prior n value by 10, compute hypothetical posterior $f_{\beta N}(\Pi_T)$ distribution and approximate it by the extended beta distribution. If the probability of material overstatement is still

---

[71]Felix and Grimlund, p. 35.

[72]The procedure for doing this using the extended beta is outlined in Grimlund, "A Framework for the Integration of Auditing Evidence," p. 222. The parallel logic using the three parameter gamm approximation is perhaps better explained in Felix and Grimlund, pp. 38-39.

In a direct communication with Richard Grimlund, he indicated to the researcher that the extended beta approximation is more accurate and, hence, that is the one used in the simulation.

[73]And as indicated on p 206 of chapter four, if the

greater than .05 repeat step 2. If the probability is less than .05 go to step 3. If n = 120, go to step 3.[74]

Step 3. Using the substantive sample size computed in the earlier steps, take a statistical sample of that size and revise the prior $f_{\beta N}(\Pi_T)$ using the sample results. Compute the posterior $f_{\beta N}(\Pi_T)$ and approximate it by an extended beta distribution. Using the inverse function procedure determine if the probability of material overstatement is > .05. If it is, reject the book value; otherwise accept.

An interesting aspect of the Felix and Grimlund model is the claim by the authors (and other accounting researchers such as Barry Cushing) that the Bayesian approach allows one to control for risks of Type II errors similar to the classical hypothesis testing approach. Although this is frequently possible when no prior information is allowed, the interpretations of the two approaches are different (the Bayesian assumes the unknown parameter value is a variable whereas the

---

mathematical assumptions are violated, the computer program also goes to step 2. This is done to reflect the deficiencies of the Felix-Grimlund model in highly reliable environments.

[74]The process is stopped at sample size n = 120 for substantive tests because this is the largest sample size that occurs using DUS with the goals specified for the simulated auditor. (This is explained in the next section.) This appears to be a natural maximum sample size to use with the Felix-Grimlund evaluation model when the sampling method is DUS. Thus it is possible to assess whether Bayesian DUS outperforms traditional DUS, and which DUS evaluation procedure is better (i.e., TACS vs. Felix-Grimlund).

the classical statistician assumes it is fixed); and when non-diffuse priors are used even the numerical results differ.[75] Thus the Felix and Grimlund model which does not use a diffuse prior for determining the extent of substantive tests does not control for α and β risks in the classical sense. Hence, it is not obvious how to make the strategies comparable ex ante. However, it is possible to determine whether or not a particular amount (such as a book value) falls within the decision interval (whether Bayesian or non-Bayesian) and this is a feature of a strategy that is of consequence to the auditor. It is hoped that the strategies defined here are considered reasonable and important enough from the auditor's standpoint.

In simulating the Felix-Grimlund model, it is felt that a change in the application of the model is needed to make it more feasible for use in the simulated environment (and, therefore, probably for most real accounting environments). Felix and Grimlund apparently stressed the application of the model to errors in individual items of an account balance population (e.g., individual trade accounts from an accounts receivable file). Considering the high skewness these populations typically have, and the strong likelihood that errors are proportional to the item size, it is unlikely that a normal distribution assumption for the size of the individual error will be very

---

[75] See the computations on p193 of this chapter for the comparability, and Winkler, pp. 421-423.

accurate. Of course, the problem may be reduced somewhat if stratification is used with modeling of the error size distribution for each of the strata separately.[76] But then new complications are introduced by the fact a separate preposterior analysis would need to be done for each stratum and somehow integrated for the entire population. A real problem would arise because in low error rate conditions (i.e., highly reliable accounting systems) there would be a paucity of errors in many of the strata on which to make the analysis (the same old data availability issue again).

It appears these problems can be largely avoided by simply redefining the population of interest much as DUS advocates have been arguing for all along. That is, merely redefine the population χ to be the population of individual dollars making up the items and not the items themselves. By this simple convention the normal error size assumption becomes much easier to accept and the whole original analysis remains intact without having to worry about possible integration of analyses of different strata.

The only changes that need to be made in using this dollar unit approach are that the substantive tests must sample the dollars not the accounts (i.e., use DUS instead of the more conventional physical item sampling) and attributes must be sampled via dollars so that a monetary error rate per dollar can be estimated instead of monetary error rate

---

[76]However, simulation studies have shown that the improvement from using stratification for difference and ratio estimators can be only marginal if at all--the same might also be true for the Felix and Grimlund model. See the Neter-Loebecke Study p. 139.

per physical item (i.e., proportion of dollars having monetary errors, not proportion of account items having monetary errors--this necessitates use of alternative I(C) for internal control information). It appears that this is the most feasible approach in implementing the Felix-Grimlund model.

III Substantive testing methods

Besides the untested Felix-Grimlund method, only two methods used in practice are considered in the simulation. This is because prior research indicates these two--stratified mean-per-unit and DUS---are the most likely to be considered the single best general purpose substantive testing methods. (See the discussion in chapter two.)

III(A) Stratified mean-per-unit. Discussion: The Neter-Loebbecke study has shown the stratified mean-per-unit (STMPU) method to be the best performing of the substantive testing methods they simulated.[77] Hence it is used in the simulation.

By the time a strategy gets to the point of simulating a substantive testing method $\alpha = .05$, β has been set, and materiality $M = .05$ of the book value. With these specifications it is possible to simulate the STMPU method.

The number of strata to use appears to be an essentially arbitrary decision. Neter and Loebbecke examined two levels of stratification,

[77] See Neter and Loebbecke, p. 87 and p. 138.

15 and 20 strata, and concluded more research would be useful at both lower and greater degrees of stratification. Cochran indicates that there is little gain in most audit situations from using more than six strata.[78] In addition, according to Don Roberts, "In some limited empirical work, it was found that using up to about five strata can be expected to result in large savings of sample size. With more strata, the incremental savings persists but becomes appreciably smaller because a few differences of larger size than anticipated can adversely affect the sample evaluation".[79]

On the other hand, it appears that these researchers may not have considered the full implications of the number of strata for sample planning purposes. This is a very important consideration in actual practice because it turns out that for the book value population used in the simulation, planned sample sizes can vary considerably depending on the number of strata. For example, use of 10 strata results in a sample size of 564, use of 15 strata results in a sample size of 327, and use of 20 strata results in a sample size of 225. Thus for a realistic simulation (i.e., one that results in realistic audit sample sizes given the goals of the simulated auditor) one is forced to consider large numbers of strata.

Another factor which can affect the sample size is the way stratification is accomplished. Two methods are widely used in audit

---

[78]William G. Cochran, Sampling Techniques, (John Wiley & Sons, New York, 1977, Third Edition), pp. 132-133.

[79]Don Roberts, Statistical Auditing, p. 96

practice: (1) form the strata so that each contains an approximately equal recorded amount (book value), and (2) form the strata so that each has approximately the same standard deviation of recorded amounts (Neymann allocation).[80] The latter method can be an optimal method in the sense the standard deviation is minimized for the mean-per-unit estimator for a given sample size and given number of strata; however, this is only assured when there are no errors in the population. When there are errors, neither method is clearly superior for all patterns of errors.[81]

The first method of stratification is used in the simulation. There are several reasons for making this choice. First, this is a much simpler method to apply than Neymann allocation (even Neter-Loebbecke resorted to only an approximation of the Neymann allocation). Second, when there are errors in the population (i.e., differences between book values and audit values) the simpler method may actually be better. Finally, the chief effect of the differences in allocation procedure is the difference in resultant planned sample size (Neymann allocation results in a smaller planned sample size), and this difference is reduced for large numbers of strata.[82]

---

[80] Ibid, p. 99

[81] Ibid, p. 100

[82] For example, see Roberts, p. 102 for the relatively small difference (10%) in the planned sample sizes when using just four strata.

With this simpler stratification procedure it was decided to use 20 strata for the stratified mean-per-unit estimator. The reasons for this choice are (1) it results in planned sample sizes more in line with that used in other research and which appear to be used in practice, and (2) the Neter-Loebbecke Study using Neymann allocation did not result in problems of overstratification using this number of strata.[83]

Since all the simulated accounting populations have net overstatement errors it seems most logical to consider one tailed tests only. This then is assumed for setting up all the hypothesis tests. In addition, there are two possible approaches to hypothesis testing used in practice with the stratified mean-per-unit estimator: the negative approach (described in appendix VI) and the positive approach described here.

Given α, β, and materiality, M, the desired precision A can be computed as follows:

---

[83]There are two aspects to the problem of overstratification in auditing: without errors, overstratification means there is barely any gain in precision of the estimator with the additional strata; with errors, overstratification can result in confidence declines (for example, see Neter and Loebbecke pp. 87-88) and some precision problems (Neter and Loebbecke, p. 66). Overall though, the two levels of stratification used by Neter and Loebbecke were comparable in performing with the reliability of the one sided confidence interval being somewhat higher for 20 strata in population three--the population used in the dissertation. See Neter and Loebbecke, pp. 88-89.

$$A = \frac{z_\alpha M}{z_\alpha + z_\beta}$$

where z_x is the normal table value which includes an area of .5 - X. (X = α, β).[84]

The 20 strata are formed so that they each contain approximately an equal recorded amount of dollar value. Once the strata boundaries are determined and the population items are divided into strata, the sample size can be determined using the following formula:

$$n = \frac{z_\alpha^{\;2}\, Y \Sigma\, N_i^{\;2} \frac{\sigma_{Y_i}^{\;2}}{Y_i}}{A^2 + z_\alpha^{\;2}\, \Sigma N_i \sigma_{Y_i}^{\;2}}$$

where Y = book value of the population

$\sigma_{Y_i}^2$ = the variance of the book values in the i'th stratum

Y_i = book value of the i'th stratum

N_i = number of items in the i'th stratum

z_α = normal table value which includes an area of .5 - α.[85]

The sample size n is then allocated in proportion to the total recorded amount in each stratum. Once the sample size for each stratum is so determined a systematic random sample is taken within each stratum. This should result in random sampling without replacement within a stratum since the items in the population are randomly re-arranged before stratification. (Systematic sampling is a widely

---

[84] Don Roberts, Statistical Auditing, p. 41.

[85] Ibid, p. 102.

accepted procedure in auditing even without prior randomization of the accounting population values.)[86]

Once the sample results have been obtained (i.e., the audit value is obtained for each of the sampled items), the mean and variance for the sample audit values from the i'th stratum are:

$$\bar{X}_i = \frac{\sum_{h=1}^{n_i} X_{ih}}{n_i} \qquad S_i^2 = \frac{\sum_{h=1}^{n_i} (X_{ih} - \bar{X}_i)}{n_i - 1}$$

The mean-per-unit estimator for the population total audit value with stratified random sampling then is:

$\hat{X} = \sum_{i=1}^{20} N_i \bar{X}_i$ and the estimated variance of this estimator is:

$$S^2(\hat{X}) = \sum_{i=1}^{20} N_i^2 (1 - \frac{n_i}{N_i}) \frac{S_i^2}{n_i}$$

This is then used to compute the achieved precision,

$$A' = z_\alpha \sqrt{\sum N_i (N_i - n_i) \frac{S_i^2}{n_i}}$$ [87]

In order to control for the β risk at the planned level (because it is the more serious risk), the achieved precision A' is frequently adjusted based upon the planned precision A, the achieved precision,

---

[86] See, for example, the Haskins and Sells manual, frames 1-64 and 1-67.

[87] Don Roberts, Statistical Auditing, p. 103.

and the materiality M.[88]

$$A'' = A' + M(\frac{A-A'}{A})$$

The one sided hypothesis test for overstatement then is:

$$\hat{X} \leq UB = \hat{X} + A''$$

where X = total audit value

UB = upper confidence limit for the specified one sided confidence (.95)

$\hat{X}$ = estimate of total audit value

A" = adjusted precision

Y = total book value (which is assumed known)

The decision rule under the positive approach is thus: if $Y < \hat{X} + A''$, then accept the book value; otherwise reject it.

Using the above rule, statistical auditing theory predict that the risk of committing a Type I error is no greater than α when there are no errors, the risk of making a Type II error as a result of applying the audit strategy is no greater than .05, where materiality = M. However, these risks are all nominal values only. The simulation measures how close the actual risks are to these nominal values.

---

[88] This is a common procedure used in practice. For example, see Roberts, pp. 43-45, or the Ernst and Ernst manual. It turns out that this procedure is required to hold down both α and β risks. In fact, as shown in chapter five, the negative approach and the positive approach with adjustment are statistically equivalent. Also, see A.D. Teitlebaum and C.F. Robinson, "A Reply," Journal of Accounting Research, 1975 Supplement, pp. 95-97.

III(B) Dollar-unit sampling. Discussion: Dollar-unit sampling (DUS) appears to be the other single most supported substantive testing method.[89]

An interesting feature of the method as used in practice (and which it has in common with the Felix-Grimlund model) is that there is no explicit control for the α risk. This is in sharp contrast to the classical STMPU estimator described above. This can be traced to the general philosophy not only of DUS advocates but the auditing profession in general that the overwhelmingly most important risk to control for is the risk of making a Type II error.[90] The argument between using STMPU and DUS appears to be reducible to the relative degree of concern for the risk of making a Type I error. Some of the critics of the DUS approach argue that DUS completely ignores the risk of making the Type I error.[91] However, a review of the DUS literature and in particular the publications of firms using the DUS approach makes clear that risks of Type I errors are indirectly controlled for.[92] This is achieved by computing a sample size based

---

[89]See chapter two of the dissertation; or A.D. Teitlebaum and C.F. Robinson, "The Real Risks in Audit Sampling," Journal of Accounting Research, Supplement 1975, pp. 70-98; and Teitlebaum, pp. 26-28.

[90]For example, see reference in footnote 89; Roberts, p. 41, recognizes this is the more serious error; and SAS No. 1 Sec. 320 refers almost exclusively to the risk of Type II error.

[91]For example, see Don Roberts, "Discussion of the Real Risks in Audit Sampling," Journal of Accounting Research Supplement 1975, pp. 92-94; Robert Kaplan "Sample Size Computations for Dollar-Unit Sampling," Journal of Accounting Research Supplement 1975, pp. 126-133; John Neter, James K. Loebbecke, James L. Goddfellow, "Some Perspectives on CAV Sampling Plans," CA Magazine, Nov. 1974, pp. 25-26.

[92]See for example Teitlebaum, Leslie, and Anderson, "An Analysis

on a projection of the material upper error limit and making a "reasonable" allowance for expected sample errors to be found. More specifically "the auditor merely subtracts from total dollar materiality the most likely error value anticipated for the audit as a whole; the resulting 'tolerable precision' then determines the dollar-unit sample size required."[93] That the primary means of controlling the α risk thru DUS is by adjusting for expected errors is further made evident in the following extract:

> "With reference to a last point, in his concern that the dollar-unit sample may turn up a single error, thus causing the upper-error limit to exceed the threshold of materiality, and subsequently causing the auditor to extend his auditing procedures far too often, Roberts clearly has simply been involved in poor planning; if there is a reasonable likelihood that a few errors (small or large) may be found in the sample (as indicated by internal control, other audit evidence, etc...), then clearly the auditor has no business taking a discovery sample. The appropriate sample size should be one which, when evaluated on containing a few errors will nevertheless not exceed the threshold of materiality. It is unreasonable to expect the sampling plan to work well if no sample planning is involved"[94]

Hence, a simulation of a strategy using DUS should incorporate these features if a valid DUS method is intended.[95] Fortunately, the Clarkeson Gordon & Co's statistical sampling manual is fairly explicit about the way sample size planning is implemented using DUS. Basically, this involves reducing the threshold of materiality (.05 in the

---

of Recent Commentary on Dollar Unit Sampling in Auditing," pp. 18-20 and pp. 27-35.

[93]Ibid. p. 86.

[94]Teitlebaum and Robinson, pp. 96-97.

[95]In referring to the Neter-Loebbecke Study, Anderson and Leslie

simulation) by the expected monetary error rate and precision gap widening (increasing precision when errors are discovered--according to the Clarkeson, Gordon & Co. manual, a very rough rule of thumb of precision gap widening amounts to 1/2 of the most likely errors, see p. 150 of their manual). However, apparently it is firm policy to not reduce basic precision to less than 1/2 of materiality. That is, if the most likely error and precision gap widening add up to something greater than 1/2 of materiality (1/2 x .05 = .025), basic precision is automatically left at .025. (See p. 148 of the manual). Thus the maximum sample size for DUS will occur with no internal control reliance and a basic precision of .025 i.e., $n = \frac{B}{p} = \frac{3.0}{.025} = 120$ (Note a pure discovery sample size would be $\frac{3.0}{.05} = 60$), and the minimum sample size will be about $n = \frac{B}{p} = \frac{.693}{.04} = 18$, where B = 3.0 is the reliability factor for a β risk of .05 and B = .693 is the factor for a β risk of .5; P = .025 is the minimum basic precision and p = .04 is an approximate maximum basic precision (based on assuming the most reliable population has an actual dollar error rate of about .006).

Note again that a α risk is essentially controlled by setting P via the expected dollar error rate, while the β risk determines the B value. (Strictly statistically speaking, B is the value such that $e^{-B} = \beta$ --this follows from the β risk associated with a Poisson

---

remarked: "While we believe the paper and study have been most valuable in identifying areas of unstated risk, it is perhaps unfortunate that the authors have not as yet studied the DUS or CMA evaluation methods as they are actually used in the field." R.J. Anderson and D.A. Leslie, "Choosing Statistical Sampling Procedures," Journal of Accounting Research Supplement 1975, p. 57.

distribution with mean = B and probability of finding K or less errors:

$$e^{-B}\sum_{j=0}^{K}\frac{B^j}{j!} = \beta.$$

For a discovery sample, K = 0, this expression reduces to $e^{-B}=\beta$).

It will be interesting to see what the actual α and β are when using this strategy.

Once the sample size n for substantive testing has been computed, a dollar unit sample is obtained by taking a systematic sample of the entire population of dollars associated with the file of records.[96]

The sample is then evaluated using the load and spread method (i.e., TACS evaluation) developed by Teitlebaum to evaluate DUS results with a minimum of conservatism.[97] To illustrate this method let B = β^{*}_{K} be the mean of the Poisson random variable X, i.e., for a given risk level β and a number of sample errors K, $P(X \leq K) = \beta$. (Note: $B = {}_{\beta}P_K$ for K errors; for a discovery sample $B = {}_{\beta}P_o$). Then assuming r overstatement errors with taintings t_i (where $t_i = \frac{\text{book value of record - audit value of record}}{\text{book value of record}}$ for each dollar i having a tainting t) and some understatement errors, h_1 the two groups of taintings are separated and evaluated separately.[98]

---

[96]Again, since the records are randomly scrambled to start with and, in addition, the monetary errors are randomly generated, a systematic sample closely approximates a proportional sample without replacement. Under these conditions a systematic sample also closely approximates the TAC selection procedure. See Anderson, p. 365.

[97]As discussed in chapter two, the TACS evaluation is designed to eliminate most of the conservatism assoicated with DUS. See Teitlebaum, pp. 12-13, pp. 27-28, and pp. 22-27 of appendix III.

[98]This part of the procedure can be found in Teitlebaum, p. 8, or

Evaluation of overstatements: order the r errors in decreasing error size $t_1 > t_2 > t_3 \ldots t_r$ and let t_o be the maximum possible tainting for undiscovered errors (in the simulation $t_o = 1$), then:

$$UEL_o = \frac{1}{n} t_o ({}_\beta P_o) = \frac{1}{n} ({}_\beta P_o)$$

UEL_1 = the greater of $UEL_o + \frac{1}{n}(t_1)$ or $\frac{1}{n} t_1 ({}_\beta P_1)$;

and in general:

UEL_i = the greater of: $UEL_{i-1} + \frac{1}{n} t_i$ or $\frac{1}{n} t_i ({}_\beta P_i)$.

Understatement: $MLE_h = (\sum_{j=1}^{h} t_i)/n$

then the net UEL = $UEL_r + MLE_h$.[99]

The decision rule is thus to reject the population if M = .05 > net UEL, and to accept it otherwise. This appears most consistent with the rule used for the stratified MPU test.[100]

III(C) Felix-Grimlund estimator. The last of the substantive testing methods considered in the simulation is the unproven but potentially useful Felix-Grimlund method outlined on pp. 200-208. In particular, step 3 on p. 208 specifies the statistical decision rule of the simulated auditor using this substantive testing method.

---

the Clarkeson, Gordon & Co. manual, p. 85.

[99]These rules can be found in Teitlebaum, p. 23, and the Clarkeson, Gordon & Co. manual, p. 195.

[100]Although this is not readily apparent when the positive approach is used with stratified mean-per-unit, it is apparent when the equivalent negative approach is used. See appendix VI.

4.4 The sixteen audit sampling strategies specified

Perhaps it is best to now list all the strategies that are simulated since not all combinations of the alternatives at each of the three stages are relevant or equally interesting. The following then is a list of the proposed strategies. The three alphabetical characters indicate the alternatives at each stage of the strategy where the positioning indicates the stage; e.g., strategy CBA is the strategy using alternative C of stage I (amount of internal control information) alternative B of stage II (linkage rule) and alternative A of stage III (substantive testing method).

Strategy 1: AAA: The auditor has perfect information about the reliability of the accounting system (the omniscience case), and he uses the linkage rule II(A) (described on pp. 181 of chapter four) to determine the extent of substantive testing using the stratified mean-per-unit substantive testing procedure.

A clarification should perhaps be made about what is considered an omniscient auditor. In a purely Bayesian sense a truly omniscient auditor would not have to do any substantive testing. He would know with certitude the condition of the accounting environment. However, an omniscient auditor in the simulation is omniscient only with respect to the internal control system—there always exists some possibility of other sources of error such as management override or collusive fraud. In addition, due to institutional factors the auditor is required to do a certain amount of substantive testing for certain items.[101] These qualitative factors are not reflected in the Bayesian Felix-Grimlund model and, as a consequence, this model does forego the use of substantive tests in certain situations as is shown in chapter five.

The non-Bayesian strategies, on the other hand, do reflect these institutional and qualitative factors and therefore always conduct a certain amount of substantive testing even for the omniscient case. However, the amount of such testing can vary even in the omniscient case depending on the linkage rule used. The reason for this is to capture the effect of the individual linkage rule. It can be argued that an omniscient auditor would always allow maximum reliance possible for the linkage rule used as long as he is aware that system reliability is greater than .9; and this situation is effectively captured for all linkage and all substantive testing methods. However, it is also of interest to consider the omniscient state as representing somewhat of a limiting case for use with each linkage rule. Thus, for example, an omniscient auditor using linkage II(A), which recognizes varying grades of internal control quality, would probably provide a better comparison with a less informed auditor using the same linkage rule if it were assumed the omniscient auditor strictly followed reliance according to these grades just as the less informed auditor. It is in this sense that a simulated auditor is considered omniscient in the dissertation. That is, the omniscient auditor is assumed to follow the

[101]See chapter two and chapter three for a discussion of these factors.

linkage rule strictly, even though there is no uncertainty associated with his assessment of system reliability. Again, the reason for this is to obtain better comparability to a more realistic strategy where there is uncertainty about the reliability of internal controls.[102] The omniscient strategies thus represent the maximum or limiting value of internal control information when a particular linkage rule and substantive test method are used.

Strategy 1, therefore, assesses the maximum value of internal control information when linkage rule II(A) is used with stratified mean-per-unit estimation. In this case the linkage rule is reduced to the following: for E1 and E2, set $\beta = .5$; for E3 set $\beta = .10$; and for E4 and E5, set $\beta = .05$.[103]

Strategy 2: ABA: Same as strategy 1 except linkage rule II(B) is used. The reason for having strategy 2 is to assess the maximum value of internal control information when linkage rule II(B) is used with stratified mean-per-unit estimation. Since linkage rule II(B) (described on pp. 196) is based on the likelihood that the monetary error

---

[102]Such uncertainty is introduced by sampling and nonsampling errors. The latter includes judgmental errors.

[103]As mentioned on pp. 181 of chapter four, an attempt has been made to reduce but not eliminate the inherent conservatism of basing reliance on grades of internal control. This conservatism is now made evident by the fact that in E3 reliance can still be complete because the error condition is still mathematically immaterial, yet β is only set to .10 instead of .5. On the other hand, this conservatism proves to be very useful when there is uncertainty associated with system reliability because it automatically limits the rise in unwarranted reliance that takes place with less than perfect knowledge of internal controls.

Finally, it ought to be noted that materiality itself is a vague rate is material, this reduces to a two state representation with omniscience. That is, since E1, E2, and E3 by definition have less than the exactly material error rate, $\beta = .5$ for these three environments. Similarly, because E4 and E5 have material errors, $\beta = .5$.[104] Thus the conservatism of strategy 1 is completely eliminated when using strategy 2. In fact this is the least conservative of all the non-Bayesian audit strategies and therefore it appears superior when the auditor is omniscient.

However, with less than perfect information, unwarranted reliance errors, which ultimately lead to Type II errors, cause the most problems with this linkage rule as is shown in chapter five.

Strategy 3: AAB: Same as strategy 1 and for the same reasons except that here DUS is used for substantive testing.

Strategy 4: AAB: Again, the interest is in obtaining an upper bound on the value of internal control information using linkage rule II(B). This strategy is the same as strategy 2 except now the substantive test method is DUS.

Note that the first four strategies are really attempting to

---

concept in auditing, and some auditors argue that auditors cannot determine materiality to within 30% to 40% of itself. See Anderson, p. 358.

In view of these uncertainties, the gradual approach represented by linkage II(A) is eminently sensible; and so, again, the omniscient strategy should be viewed more in terms of representing an upper bound on efficiency with less and less uncertainty associated with using a particular linkage rule.

[104]This arises because the omniscient auditor has a very high

ascertain the performance of the two linkage rules A and B with the two substantive test methods of A and B (i.e., stratified mean-per-unit and DUS, respectively).

Strategy 5: ACB: Perfect information on compliance error rate per dollar, and a linkage rule used by some DUS auditing firms (linkage rule II(C) is described on pp. 30-44). In this perfect information case it is assumed that for immaterial errors in the record file (i.e., for E1, E2, and E3), $\beta = .2$; and in E4 and E5, $\beta = .05$. Simulating this strategy is useful for showing the effects of eliminating the conservatism associated with the three times materiality rule for compliance deviations.[105] This linkage rule is used only with the DUS method for substantive testing.[106]

Strategy 6: ADC: This is the Felix-Grimlund model assuming a very high prior weighting ($n^* = 750$) attached to internal control information and assuming the means for all distributions used in the model equals the actual values used in the simulation of the environmental process. This high weight attached to the internal control information

---

likelihood value C = 1 in E1, E2, and E3 and a very low likelihood value C = 0 in E4 and E5. Using the linkage II(B) formula given on pp. 196 of this chapter $\beta = \frac{.05}{1-0} = .05$ in E4 and E5 and $\beta = \frac{.05}{1-1} \geq .5$ implies β is set to .5 by convention. Note that without the convention of setting the upper-bound on the value of β to .5, this linkage rule is very similar to a purely Bayesian one in the sense all substantive tests might be eliminated with sufficient information.

[105]See pp 198-200 for a description of this rule.

[106]Since random sampling without replacement is the most common form of sampling for compliance tests, it is assumed such sampling is appropriate for use with DUS for substantive tests. However, it appears highly unlikely that firms using DUS for compliance testing would not

means that all mathematical conditions of the model are satisfied.[107]

Strategy 7: BAA: This strategy uses objective information (i.e., based on compliance tests) on the internal control system via lower bounds on system reliability at 95% confidence. This strategy assesses the impact of less than perfect, and more realistic, degree of internal control information on the audit process. As discussed earlier the size of the sample for compliance tests on error rates = 150 for each attribute. The lower bound on system reliability $\hat{\underline{R}}_{.95}$ is computed using the Mann method described in Appendix IV. This $\hat{\underline{R}}_{.95}$ is then interpreted using linkage rule II(A) (described on pp. 181 of this chapter) and a planned β level is thus obtained. With β determined, α = .05, and M = .05 of the total book value, a stratified mean-per-unit test is constructed as described on pp. 214-17 of this chapter.

Strategy 8: BBA: Objective information on error rates via the sample of 150 for each of the five attributes is obtained and converted to the associated probability, C, that the system reliability, R, is greater than .90. This probability, C, is computed using the Mann method described in appendix IV. As discussed on pp. 191-96 of this chapter, C represents the reliance the auditor places on the system of internal controls. The associated β value is then computed by using

---

also use it for substantive testing. Thus a strategy using DUS for attributes and stratified mean-per-unit for substantive tests has not been considered in the simulation. See Felix and Goodfellow, p. 17.

[107]Note that this high prior weight implies a compliance test sample size of 450.--triple the objective internal control information state.

See formula on p. 202 of chapter four.

the SAS No. 1 Sec. 320B formula: $\beta = \frac{.05}{1-C}$.[108] This β value is then used along with $\alpha = .05$ and M = .05 of total book value to compute the stratified mean-per-unit test described on pp. 211-17 of this chapter.

Strategy 9: BAB: Same as strategy 7 except that one β is determined using linkage II(A), a hypothesis test based on DUS as described on pp. 60-63 of this chapter is used.

Strategy 10: BBB: Same as strategy 8 except that once β is determined using linkage II(B), a DUS hypothesis test is constructed as described on pp. 218-222 of this chapter.

Note strategies 7 and 8 are used to isolate the impact of the two linkage rules II(A) and II(B) when the substantive test method is STMPU. Similarly, strategies 9 and 10 are used to accomplish the same purpose when the substantive test method is DUS.

Strategy 11: CCB: This is the strategy using the linkage rule unique to DUS where the three times materiality smoke/fire ratio linkage rule is used (see pp 198-200 of this chapter). A 150 sample size compliance test is taken for each of the five attributes and a system compliance error rate (SCER) upper error limit is estimated using the Mann method (i.e., the upper error limit on the proportion of dollars having at least one compliance deviation is estimated). This estimate is obtained by using DUS attribute sampling described on p. 173.

Linkage rule II(C) thus reduces to the following rule in the simulation: if the upper error limit on SCER $\geq$ three times materiality $= 3 \times .05 = .15$, the β risk for substantive tests is set to .05 (i.e.,

---

[108] See pp. 196 and footnote 104 of this chapter to see how this linkage formula is obtained from SAS No. 1.

let β = .05); otherwise set β = .2. With these levels of β so set, a DUS test is constructed as described on pp. 218-222.

Strategy 12: CDC: This is the Bayesian Felix-Grimlund model with the prior distribution for substantive tests established by compliance testing as described on pp. 200-11. When the prior is based on compliance test information obtained from a sample size of 150 for each of the attributes the prior information has a weight of n* = 250 equivalent to a substantive test of the financial output of the system as described on p. 205. This and other Felix-Grimlund strategies are more completely described on pp. 200-11 of this chapter and in appendix V.

Strategy 13: D-A: No reliance on internal controls attempted, hence no linkage rule and no value to internal control information because β is automatically set to .05, which is the combined risk level. The reason for having this strategy is to serve as a benchmark for the others which attempt to rely on internal control information to some extent. If this "informationless" strategy is not outperformed significantly by those using internal control information, then auditors should begin to seriously reassess the importance of internal control information or at least the strategies that are being simulated here. The substantive test method used for this strategy is STMPU.

Strategy 14: D-B: No reliance is planned on internal controls, hence no linkage rule, β = .05 every time. Now, however, the DUS test is used as described on pp. 222. Again, this strategy serves as a bench mark representing the "informationless" state using DUS.

Strategy 15: D-C: Again, no internal control information is available and hence a constant sample size is used for all five environments. Since DUS is the basis of the selection procedure, it appears most natural to use the same sample size as the informationless DUS strategy 14.[109]

Strategy 16: CAB: This is the strategy that attempts to introduce judgmental error to the audit process in hopes of simulating the effect of judgmental error in integrating internal control information with substantive testing. (See pp. 174-6) The linkage rule and substantive test alternatives indicated prove to be the best performing of the objective strategies simulated (i.e., strategy 9: BAB, outperformed all other strategies obtaining information about internal controls thru compliance testing--strategies 7-12 inclusive). The implications of simulating this strategy are more completely described in chapter five.

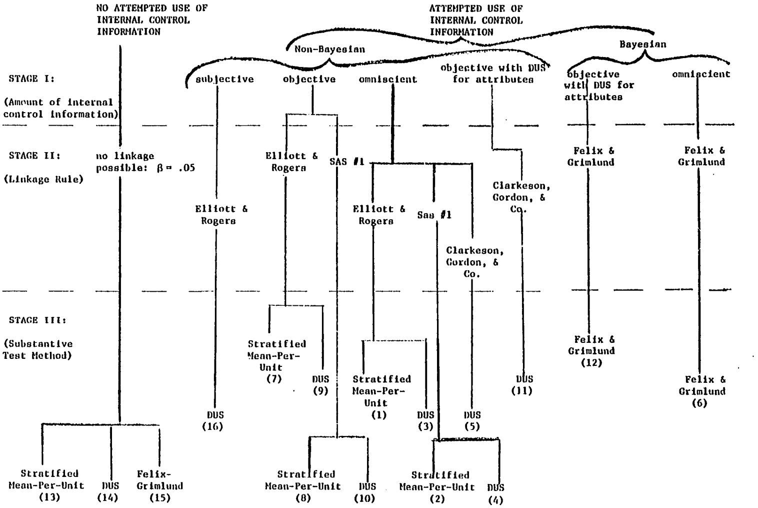
To facilitate comparisons of the audit sampling strategies, these strategies are diagramed in figure 6 given on the next page.

Once these strategies have been simulated, it is possible by contrasting the performances of the strategies to determine the impact of

---

[109]The reasoning for limiting the maximum sample size to the maximum used by the more traditional DUS procedure is that if the Felix-Grimlund model cannot outperform the more conventional DUS model under comparable conditions, there is little reason to consider the Felix-Grimlund model further.

FIGURE 6
DIAGRAM OF SAMPLING STRATEGIES*



*Strategy numbers are given in parentheses at bottom of each strategy.

(1) different levels of internal information: (a) perfect information, (b) objective information via compliance tests, (c) subjective information and (d) no information; (2) different linkage rules; and (3) different substantive tests. How the performance evaluation is to be made and the overall experimental design are discussed next.[110]

4.5 Rules for comparing strategy performance

As discussed in chapter one, the primary measures of performance of a sampling strategy as far as auditor decision making is concerned, are the substantive test sample size distribution, and the α and combined risks associated with a strategy. The environments have been constructed to facilitate comparison of these measures since, essentially, these are the most important measures of a strategy that a researcher can provide to help the auditor decide on a strategy.

However, in addition to these measures for each strategy and environment combination, it may be useful to average these measures to summary measures relating to a strategy only. This is possible by making assumptions about the frequency with which the environments occur. The five environments represent a useful relevant range in terms of percentage of material dollar error over which most actual environments can be expected to fall. Each environment can thus be thought of as representing a certain range of materiality of error. Thus E1 can be thought of as representing environments having errors

[110] That these strategies include the strategies presently of most interest to the auditing profession should be evident on reading chapter two.

up to, say, 1/3 materiality, E2 up to 2/3 materiality, and E3 between 2/3 and up to, but not including, materiality itself. Similarly, E4 can represent, say, all environments between exact materiality and 1.5 materiality, and E5 represents all environments with dollar errors greater than 1.5 times the material amount.

Now, an auditor on obtaining the results of this study may find it useful to make assumptions about the relative frequency of certain environmental conditions in evaluating a sampling strategy. In particular, it may be useful to make assumptions about the relative frequencies with which error conditions in terms of materiality occur in actual accounting environments. This can have important consequences for the risks associated with a particular strategy.

To illustrate the analysis and the effect in evaluation of making different assumptions, two different assumptions about frequencies of environmental conditions are used in the study. The first assumption is based on information made available by the firm of Clarkeson, Gordon & Co. of Canada. According to it most populations have less than 1/3 a material amount of error and only about 5% of the time are material errors indicated by the audit.[111] This leads to the following probably somewhat conservative but "realistic" error frequency assumption: E1 (1/3 materiality) has a probability of occurence of .5 associated with it, E2 (1/3 to 2/3 materiality) has a probability of .3 associated

---

[111] See the Clarkeson, Gordon & Co. manual p. 164 and 178.

with it, E3 (2/3 to 3/3 materiality) has a probability of .15 associated with it, E4 (1 to 1.5 materiality) has a probability of .025 associated with it, and E5 (over 1.5 materiality) has a probability of .025. With these assumptions it is now possible to approximate an average α and combined risk, and substantive test sample size associated with an audit strategy.

Another assumption about error size frequencies is one that is clearly conservative given the experiences quoted in footnote 111 and so is probably conservative for most audit firms: assume all five environments have an equal likelihood or probability of occurring. This means that each environment has a probability of .2 of occurring and this is clearly conservative because it implies that 60% of all audits (i.e., E3, E4, E5) encounter material amounts of error, or close to it. This equal weighting should thus provide a conservative assessment of expected α, combined risk, and substantive test sample size associated with a strategy.

If the α and combined risks are not significantly different, the weighting assumptions used by an auditor should not be expected to change the ranking of a strategy. This is because in such a situation the substantive test sample size is the primary determinant in evaluating a strategy. It turns out that for many strategies this sample size is consistently smaller than or equal to that for another strategy. Thus no matter how the sample sizes may be weighted one strategy would consistently outrank the other. For example, assume that the only difference between strategy A and strategy B is that strategy A uses internal control information but strategy B does not (and hence uses constant sample sizes for all environments). Assume the following substantive sample sizes are obtained:

	E1	E2	E3	E4	E5
Strategy A	50	200	295	300	300
Strategy B	300	300	300	300	300

Assuming that actual α and combined risks are not significantly different and using the conservative weighting scheme, strategy A has an average substantive sample size of 229 while strategy B has one of 300, i.e., A is preferred to B. Using the more realistic weighting scheme above, strategy A has an average sample size of $.5(50) + .3(200) + .15(295) + .025(300) + .025(300) = 25 + 60 + 44.25 + 15 = 144.25$, while strategy B is still 300, i.e., A is still preferred to B. In fact strategy A is always preferred to strategy B as long as any sample size for A is not greater than the sample size for B (which is the case in this situation). Thus the impact of internal control information can be assessed in a relatively straightforward manner in the present framework.

If the sampling risks do differ significantly between strategies then the same weighting scheme for the sampling risks may still be useful. This depends on how the risks vary with different amounts of error in the accounting environment. Nevertheless, no matter how the strategies are evaluated, it is evident that the measures obtained from the simulation are necessary inputs to such an evaluation and this then is a major contribution of the dissertation.

In order to make tentative evaluations in the dissertation, it is necessary to specify the rules for determining when average substantive test sample sizes associated with a given strategy are significantly different from that of other strategies; and when the actual sampling risks associated with a strategy are significantly different. In the latter case, it is proposed that rule for deciding when two sampling risks are essentially equal is to consider them so when the two actual risks are less than 2.5 percent points apart. This is the same rule used by Neter and Loebbecke, who considered that the reliability of the nominal confidence coefficient (in estimating the actual proportion of correct decision intervals) was "high" if nominal reliability was within 2.5 percent points of the actual reliability.[112] This rule is considered reasonable for the dissertation as well.

Since the substantive test sample sizes of a strategy are conditional on both the amount of internal control information and the linkage rule, it does not seem possible to construct a statistical test of the significance of differences among strategies. Hence, the rule used to decide when the sample sizes are considered significantly different is to consider them so when there is more than a 5% difference in average sample size.

Also as in the Neter-Loebbecke Study, it is assumed that 600 simulation runs for each strategy in each accounting environment is sufficient for the analysis and so this is the number of runs used in

---

[112] See Neter and Loebbecke, p. 128.

the dissertation.[113]

Finally, to complete specification of the performance of the strategies, the following statistics are computed for the sample estimators (for both the reliability estimate and the substantive test estimate) for the extreme sample sizes used in the simulation: mean, standard deviation, skewness, kurtosis, minimum and maximum values. These statistics and the other measures of strategy performance are presented and interpreted in chapter five along with a numerical description of the five accounting environments.

---

[113] Neter and Loebbecke, p. 8.

CHAPTER FIVE

Results of the Simulation Study

5.1 Introduction

This chapter reports the results of the simulation study. Section 5.2 provides a concise summary of the evidence on the statistical validity of the internal control hypothesis. Subsequent sections essentially provide more detailed information supporting the findings reported in section 5.2. Section 5.3 presents the environmental characteristics of the five environments. The actual values of the simulated accounting environments differ somewhat from the target values reported in chapter three due to the stochastic nature of the error generating processes, but the overall conditions are as described in the third chapter.

Next follows a section reporting in greater detail the results of simulating the audit sampling strategies defined in chapter four on the environments specified in section 5.3. This results in a rather lengthy section which in turn can be conveniently divided into subsections. The first subsection reports on the comparison of the performances of the Mann method with the crude method for estimating the lower bound on series system reliability. Then follows a report on the relative performance of DUS and stratified mean-per-unit (STMPU) estimators in the five environments. After that comes a subsection analyzing the performance of all strategies using the DUS procedure for substantive testing, followed by a subsection reporting on the performance of strategies using the STMPU estimator. It is felt this organization of strategies by substantive testing methods is more useful than other contrasts of the audit strategies. The final subsection of the third section reports on the performance of the strategies using the Felix-Grimlund model.

The last section synthesizes and summarizes the results of the study and lists the major and minor conclusions reached.

5.2 Evidence on the validity of the internal control hypothesis

Without having to initially provide many of the background details, tables 1 and 2 attempt to summarize the maximum impact of internal control information. This summarization is obtained by using the two assumptions about relative frequencies of material and immaterial errors that were discussed at the end of chapter four—realistic and conservative averaging. In essence, then, the information given in tables 1 and 2 represent extremes of the long run average performance of the substantive test methods with perfect internal control information and with no such information. Performance is measured in terms of estimated average actual α and β risks, and substantive test sample sizes associated with the test over the range of possible internal control conditions.

Upon comparing the substantive test method performances, it is clear that, whichever of the two weighting schemes is used in averaging, the impact is substantial (well beyond the 5% difference

TABLE 1

CONSERVATIVE AVERAGING OF STRATEGY PERFORMANCE STATISTICS

Strategy Substantive Test Method	Amount of Internal Control Information					
	Omniscience			None		
	risk α risk	β risk	sample size	α risk	β risk	sample size
DUS	.244664	.002666	64.8	.277998	.002666	120
STMPU	.2905	.008	130.8	.284667	.008	237
Felix-Grimlund	.187999	.014666	72.0	.219333	.014666	120

TABLE 2

REALISTIC AVERAGING OF STRATEGY PERFORMANCE STATISTICS

Strategy Substantive Test Method	Amount of Internal Control Information					
	Omniscience			None		
	α risk	β risk	sample size	α risk	β risk	sample size
DUS	.245663	.000333	32.6	.2684985	.000333	120
STMPU	.156334	.001	68.85	.288833	.001	237
Felix-Grimlund	.02350	.001833	24.0	.187999	.001833	120

considered significant in chapter four) for all substantive test methods considered. In addition, the actual α and β risks are comparable to or even significantly less than with no use of internal control information. This holds even though the sample sizes with reliance are likely to be the smallest used in actual practice.

Note that in all cases actual α and β risks are much less than the planned ones given in chapter four. This is due to the fact that neither immaterial nor material errors occur on 100% of all audits. Thus the best way to interpret the α, β, and sample size figures for each substantive test method is that they represent the long run expected value for the particular weighting scheme. Of course, other weighting schemes are possible. But, as argued in chapter four, although the degree of impact might vary, the general conclusion of a significant impact of internal control information would not be changed. The overall impression of both weighting schemes (which represent the extremes based on one audit firm's experience) is that the statistical validity of the internal control hypothesis is strongly supported. That is, in all cases the auditor is far better off in terms of substantive test sample size and, sometimes, reduced sampling risks when reliable internal control information is available.[1] However, there remains the very practical question

[1]It should be reiterated that this dissertation and discussion is only the context of assessing the potential value, but not the cost, of obtaining internal control information. Thus any conclusions about cost-benefit justification would be premature.

of whether such reliable internal control information can be made available in the "real world." This and other detailed issues pertaining to the hypothesis's validity are considered in the remainder of the chapter.

5.3 Simulation of the five accounting environments

Book value distribution

The five environments are all based on the same book value distribution. This distribution is taken from population three of the Neter-Loebbecke study and was chosen primarily because of its source--accounts receivable from a medium-sized manufacturer--and the fact it is one of the more highly skewed of the Neter-Loebbecke populations. This population should therefore not only be representative of many that auditors actually encounter in practice but also useful for testing whether the high skewness causes problems for the major statistical estimators. Table 3 presents the major characteristics of the book values (BV), audit values (AV), and dollar errors (BV-AV); and figure 7 presents a frequency polygon of the distribution of the book amounts.

Attributes and Reliability

The following notation is useful in further describing the environments.

CER = compliance error rate per record = proportion of records having a particular compliance error

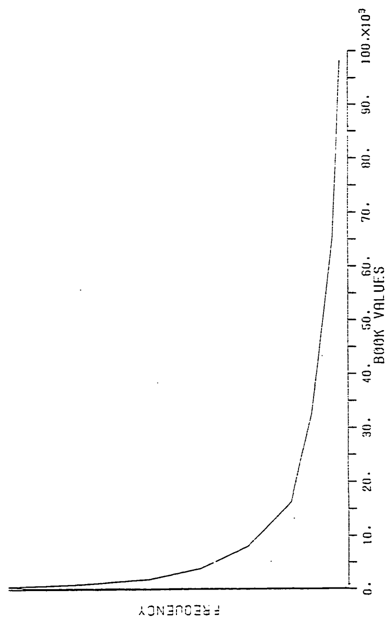
CE = number of compliance errors

TABLE 3

KEY ENVIRONMENTAL CHARACTERISTICS

	Environments				
	E1	E2	E3	E4	E5
Total audit value	13634358.70	13353290.78	12998901.89	12982816.49	12643134.33
Distribution of BV					
Mean	1946.	1946.	1946.	1946.	1946.
Standard deviation	7022.	7022.	7022.	7022.	7022.
Skewness	7.94	7.94	7.94	7.94	7.94
Kurtosis	78.17	78.17	78.17	78.17	78.17
Maximum	98163.	98163.	98163.	98163.	98163.
Minimum	0.	0.	0.	0.	0.
Distribution of AV					
Mean	1941.	1901.	1850.	1848.	1799.
Standard deviation	7017.	6881.	6740.	6739.	6618.
Skewness	7.96	7.90	7.96	7.97	8.07
Kurtosis	78.41	77.29	78.74	78.79	81.07
Maximum	98163.	98163.	98163.	98163.	98163.
Minimum	0.	0.	0.	0.	0.
Distribution of BV - AV					
Mean	5.	45.	96.	98.	146.
Standard deviation	180.	780.	1255.	1259.	1513.
Skewness	60.56	36.77	32.92	32.68	24.43
Kurtosis	4243.41	1613.35	1424.34	1409.32	797.47
Maximum	13284.	40973.	66904.	66904.	66904.
Minimum	-99.	-114.	-151.	-194.	-925.

FIG. 7: FREQUENCY POLYGON OF POPULATION 3 BOOK VALUES



CERD = compliance error rate per dollar = proportion of dollars having a particular compliance error (i.e., relating to k_1 or k_2 etc., but not the entire system of controls)

SCERD = system compliance error rate per dollar = proportion of dollars having any kind of compliance error associated with it

SCERD = $(1 - CERD_1)(1-CERD_2)(1-CERD_3)(1-CERD_4)(1-CERD_5)$

MER = monetary error rate = proportion of the records having a dollar error (either overstatement or understatement)

MERD = monetary error rate per dollar = proportion of dollars having a dollar error

ME = number of monetary errors = number of records having a dollar error (either overstatement or understatement)

AV = total audit value of population

BV = total book value of population

DER = dollar error rate = $\frac{BV - AV}{BV}$

Note: A net overstatement error means that BV - AV is positive, means that DER is positive. All environments have 7026 records and BV of $13,671,503. All environments have understatement as well as overstatement errors. All system compliance error rates or monetary error rates are the unreliabilities, 1 - R, associated with a particular reliability concept, R.

The beta distribution $F(\theta)$ used to generate the amount of dollar error is an extended beta with the following parameter values: $p = 3$, $q = 9$, $a = 0$, $b = 2$. This results in an expected error $\frac{BV - AV}{BV}$ of .5

per book value dollar given a monetary error has occurred, and a variance of .057692. This beta distribution provides the link between DER and CER or MER as explained in chapter three.

Due to the stochastic nature of the generating processes in the simulation, there are some differences between the target CER's, MER's, and DER's as given in chapter four and what was developed in the simulation. To get more exact values would have required much more computer time. Overall, though, the environmental characteristics are as intended and provide the measures of strategy performance necessitated by the goals of the dissertation.

The various error rates associated with each environment are listed in table 4. Both the targeted (as listed in chapter four) and the attained error rates are listed.

Audit value characteristics

The five environments result in five different audit value distributions varying in the amount of difference between total audit value and total book value (which is constant for all five environments). It should be noted that in actual practice the total audit value is something that is rarely if ever known with certainty. That is why there have been no published examples of audit value distributions. However, one can reasonably expect a close correlation between audit and book values in most situations (this is the implied assumption in stratifying on the basis of book values and in DUS of book values), and so the simulation is constructed on this assumption. Thus the characteristics of table 3 are of primary interest in showing the

TABLE 4

OTHER ENVIRONMENTAL CHARACTERISTICS

Characteristic	Environment				
	E1	E2	E3	E4	E5
BV	13,671,503	13,671,503	13,671,503	13,671,503	13,671,503
AV	13,634,358.64	13,353,290.72	12,998,901.83	12,982,816.43	12,643,134.25
Target CER	.006	.03	.058	.063	.096
Actual CER for k1	.000551	.029747	.059066	.062198	.096926
Actual CER for k2	.007259	.031028	.063479	.065471	.102904
Actual CER for k3	.005266	.028466	.058212	.061771	.088955
Actual CER for k4	.006120	.029320	.060917	.065186	.098349
Actual CER for k5	.006405	.029604	.056504	.058639	.092656
Target MER	.01	.05	.09	.10	.15
Theoretical MER based on actual CER's	.010159	.048422	.095520	.100152	.150027
Actual MER	.008682	.048819	.096641	.101196	.150157
Actual MERD	.005800	.049221	.100381	.102684	.150045
Actual ME	61	343	679	711	1055
Actual SCERD	.021499	.134279	.261052	.269564	.366192
Target DER	.005	.025	.0491973	.05	.075
Actual DER	.0027169	.0232756	.0491973	.0503739	.07521999

kinds of audit value distributions this and other assumptions given in chapter three give rise to. Figure 8 gives the frequency polygon for the audit values of E5. The reason only one environment's polygon is given is that all the audit value distributions are virtually indistinguishable from the book value distribution.

Error pattern characteristics

The difference between the book and audit value gives rise to the monetary amount or dollar error for a record (book value of record-audit value of record). The characteristics of this dollar error distribution are also given in table 3 for each environment. Figure 9 gives the frequency polygon for the E5 error distribution. Again, only one error distribution graph is provided because the other distributions are very similar. (In fact, they show even less detail for the same scale because there are fewer errors.)

This completes the description of the five accounting environments used in the simulation. These characteristics follow from the assumptions given in chapter three and properties reported by Ramage, Kniger, and Spero. Thus the simulated environments appear to be representative of characteristics of actual accounts receivable populations and, hence, the performance of audit strategies should not be too different from that if applied in the real world. That is, the performance of the audit strategies appears to have high external validity.

FIG. 8: FREQUENCY POLYGON OF ENVIRONMENT 5 AUDIT VALUES

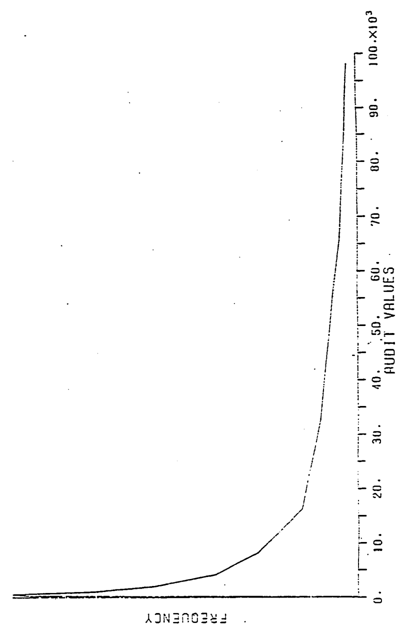
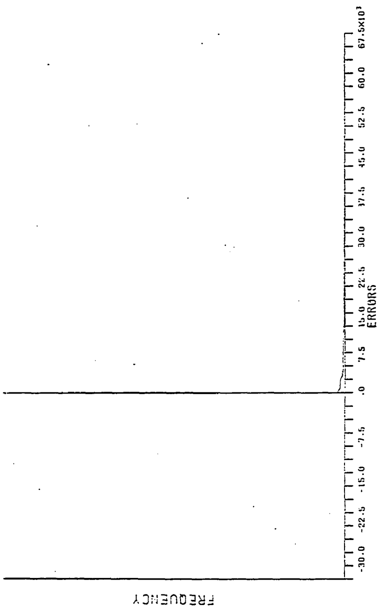


FIG. 9: FREQUENCY POLYGON OF ENVIRONMENT 5 ERRORS



5.4 Simulation of the audit sampling strategies

Comparison of Mann with crude method

The presentation of audit sampling strategy results begins with some statistics concerning the relative performance of the Mann method (described in appendix IV) and the crude method (described on p.163) for computing the lower confidence bound on system reliability. The definition of reliability used for this comparison is the probability that a record is processed without a monetary error occurring. Thus after compliance testing a maximum likelihood estimate (MLE), $\tilde{\Phi}_i$, for each attribute rate i is obtained, and this MLE estimate of system reliability is $R = (1 - \frac{\tilde{\Phi}_1}{3})(1 - \frac{\tilde{\Phi}_2}{3})(1 - \frac{\tilde{\Phi}_3}{3})(1 - \frac{\tilde{\Phi}_4}{3})(1 - \frac{\tilde{\Phi}_5}{3})$.

The same MLE estimate for R is used under both methods, it is in the computation of the lower bounds that the two methods diverge. Table 5 lists the results of ten estimates using these two methods for E1 (R = .991318), the most reliable environment.

A review of table 5 indicates that the Mann method results in a much more precise bound than the crude method. The conservatism of the crude bound is evident by the fact that not a single bound is over .95; hence, even for the most reliable environment there would not be any maximum reliance using linkage rule II(A). The Mann bound, however, would allow maximum reliance for all the sample estimates. (Note that a lower bound on system reliability is equivalent to an upper bound on the system error rate--both bounds use the same sample information.)

The conservatism of the crude method is less evident when the C value is computed but this is because E1 is the most reliable

TABLE 5

MANN VS. CRUDE RELIABILITY ESTIMATES FOR E1
(Actual R = .991318)

SAMPLE	MLE	MANN LOWER BOUND	CRUDE LOWER BOUND	MANN C VALUE	CRUDE C VALUE
1	.9801477	.979659	.914451	1.0	.970312
2	.993348	.993060	.936919	1.0	.997214
3	.993343	.993055	.937463	1.0	.994303
4	.993343	.993055	.937463	1.0	.994303
5	.995561	.995323	.941216	1.0	.998002
6	.993348	.993060	.936919	1.0	.997214
7	.988933	.988564	.928924	1.0	.992734
8	.991126	.990794	.933984	1.0	.982733
9	.993343	.993055	.937463	1.0	.994303
10	.986726	.986323	.925476	1.0	.981182
AVG.	.990922	.990595	.933028	1.0	.99023

environment. As the reliability of the system decreases, the crude method conservatism gets worse. This is illustrated in the next table which displays the same statistics for E2 (R = .951181), the next most reliable environment with error rates approximately half that of materiality.

TABLE 6

MANN VS. CRUDE RELIABILITY ESTIMATES FOR E2
(ACTUAL R = .951181)

SAMPLE	MLE	MANN LOWER BOUND	CRUDE LOWER BOUND	CRUDE C VALUE	MANN C VALUE
1	.951968	.951235	.872982	.372755	1.0
2	.949911	.949164	.868713	.637648	1.0
3	.954132	.953413	.875992	.534937	1.0
4	.954141	.953423	.875456	.518053	1.0
5	.945651	.944879	.862836	.550510	1.0
6	.947757	.946997	.866179	.512206	1.0
7	.956319	.955616	.877990	.727536	1.0
8	.939281	.938473	.854312	.380499	1.0
9	.962708	.962052	.888816	.557790	1.0
10	.949887	.949140	.869015	.511391	1.0

Note that using linkage II(A) and the crude lower bound, there would no longer be any reliance on internal controls. Similarly, the crude C value allows maximum reliance with $\beta = \frac{.05}{1-6} = \frac{.05}{1-.727536} = .1835$ and minimum $\beta = \frac{.05}{1-.372755} = .0797$. By contrast, using the Mann C value allows full reliance ($\beta = \frac{.05}{1-1} > .5$ implies $\beta = .5$) for all the sample results using linkage rule II(C). It is thus evident that using the Mann method tends to maximize the impact of internal control information and for this reason the Mann method is used in the simulation.

However, there is a drawback to using the Mann method for the computation of system bounds and C values, and this is the fact that the method results in misstated confidence associated with the bounds. In particular the nominal (95%) confidence level misstates the actual confidence level which appears to be more on the order of 50%. This appears to arise because of the very precise bounds that are computed (note the extremely small difference between the MLE and the Mann lower bound on reliability). This bias (apparently due to the approximateness of the procedure for these conditions) can lead to excessive unwarranted reliance on the internal controls as the compliance error rates become material. This is particularly true of the Mann C value which, because of the precision of the Mann bound, tends to take the extreme values of one and zero for even the smallest sampling variation. The result is that there is frequent complete reliance on internal controls (C = 1) even when compliance error rates are at the material level. This is illustrated in the next three tables.

It is evident from table 8 that the Mann C value is overoptimistic around the exactly material unreliability, 1-R, level. This arises because the confidence bound is too high for the stated level of confidence (e.g., actual confidence of 60% vs. nominal confidence of 95%

TABLE 7

MANN VS. CRUDE RELIABILITY ESTIMATES ON E3
(ACTUAL R = .903359)

SAMPLE	MLE	95% Confidence MANN LOWER BOUND	95% Confidence CRUDE LOWER BOUND	MANN C VALUE	CRUDE C VALUE
1	.887584	.886627	.788094	0.0	.000757
2	.907937	.907010	.813846	1.0	.012227
3	.903898	.902963	.808147	1.0	.015849
4	.905947	.905026	.810684	1.0	.022886
5	.918301	.917405	.826852	1.0	.057279
6	.901755	.900815	.806079	.998426	.003629
7	.914193	.913283	.821196	1.0	.055545
8	.877461	.876503	.7776104	0.0	.000044
9	.926682	.925802	.837387	1.0	.169803
10	.928755	.927898	.840227	1.0	.189127

TABLE 8

MANN VS. CRUDE ESTIMATES ON E4
(ACTUAL R = .898804)

SAMPLE	MLE	95% Confidence MANN LOWER BOUND	95% Confidence CRUDE LOWER BOUND	MANN C VALUE	CRUDE C VALUE
1	.920383	.919494	.829513	1.0	.073901
2	.914019	.913109	.822406	1.0	.007055
3	.883600	.882638	.783018	0.0	.000610
4	.903889	.902954	.808207	1.0	.014111
5	.893686	.892736	.795505	0.0	.003020
6	.901774	.900834	.806021	.998576	.004607
7	.891618	.890664	.793121	0.0	.001320
8	.891429	.890475	.794198	0.0	.000131
9	.883568	.882610	.783141	0.0	.000428
10	.889606	.888651	.790616	0.0	.001084

TABLE 9

MANN VS. CRUDE ESTIMATES ON E5
(ACTUAL R = .849843)

SAMPLE	MLE	95% Confidence MANN LOWER BOUND	95% Confidence CRUDE LOWER BOUND	MANN C VALUE	CRUDE C VALUE
1	.871419	.870466	.769202	.00	.00
2	.817256	.816511	.705498	.00	.00
3	.842104	.841223	.733975	.00	.00
4	.836118	.835261	.727575	.00	.00
5	.861619	.860680	.756918	.00	.00
6	.855616	.854691	.750417	.00	.00
7	.830551	.829724	.720670	.00	.00
8	.859323	.858386	.755167	.00	.00
9	.847912	.847009	.740778	.00	.00
10	.857730	.856800	.752140	.00	.00

for the ten samples). This thus reflects a deficiency of present theory since the Mann approximately optimal bound does not prove to be very optimal.[2]

What should be the auditor's response in the face of this shortcoming? Well, it certainly does not appear appropriate to discard the Mann method in favor of the crude method considering that the latter method results in such a small reliance on internal controls (compared to the omniscient case). Instead, a response consistent to that

[2]Considering that other methods for computing series system reliability appear to perform no better (See Winterbottom and Mann, Schaeffer and Singpurwalla), it appears all available theories have weaknesses.

proposed by Roberts appears more appropriate. Roberts argues that when internal controls are to be relied upon, the auditor should allow for additional sampling risk for his substantive test.[3] This is because the compliance testing can result in unwarranted reliance on internal controls--the auditor incorrectly decides that the procedures are being followed satisfactorily when, in fact, compliance deviations are more numerous than satisfactory. Hence, the auditor erroneously reduces the extent of substantive tests thus increasing the risk of missing a material amount of monetary error. That is, the combined risk of the audit is increased beyond the nominal level. The additional sampling risk is allowed for by, essentially, setting the nominal level below that of the actual risk level the auditor is willing to incur. For example, if the auditor is willing to incur a combined risk of .06 he may set the nominal level at .05, thus automatically adding a cushion for the risk of unwarranted reliance.[4]

The same result can also be effected by reducing the risk of Type II error for the compliance test. In the case of the C value computation, this can be achieved by setting the R value to be higher than what the auditor considers to result in an exactly material amount of dollar error. Thus, referring to appendix IV, p. 394, where the Mann method formulas for computing C are presented; by resetting R to go from .9 (exact materiality) to, say, .92 the risk of unwarranted

[3]Roberts, pp. 144-146.

[4]Based on Roberts' numerical example, p. 145.

reliance can be considerably reduced. This is what is done, then, in simulating the auditor's actions when linkage rule II(B) is used with the Mann C value:

C = confidence level associated with a z value of $(\frac{2}{9\nu})^{-1/2}[[\frac{\ln(.92)}{-ms}]^{1/3} + \frac{2}{9\nu} - 1]$. See p.392 of appendix IV for a definition of the variables. Again, the justification for this is twofold: (1) to reduce the increased risks of making a Type II error from using imperfect internal control information; and (2) to compensate for the bias of methods presently available for estimating series system reliability from component data.

This adjustment is made only for the computation of the C value because this C value is used with the linkage that eliminates all conservatism in interpreting the compliance test results. (See discussion in footnote 36 of chapter four.) For all other linkage rules and reliability calculations, the Mann method is used as indicated in appendix IV because the other linkages (i.e., linkages II(A) and II(C)) have enough inherent conservatism to nullify the effects of using the excessively high 95% confidence lower bounds of the Mann method. Thus the Mann method appears to be suitable for direct use in auditing when interpretations are made on the reliability lower bound estimates (or, equivalently, on the upper error limits of the unreliability, 1-R, of the system), but needs to be adjusted for use in auditing when the C value is computed.

Using a sample of 150 for each of the five attributes and simulating 600 independent such samples, table 10 lists the statistics for

TABLE 10

STATISTICS FOR THE ESTIMATES OF RELIABILITY

	Environments				
	E1	E2	E3	E4	E5
Actual Reliability = 1-SCERD	.978501	.865721	.738948	.730436	.633808
Distribution of R1MLE*					
Mean	.9788468	.8682849	.7313540	.7243343	.6269564
Standard deviation	.0107219	.0211598	.0334757	.0354190	.0285100
Skewness	-.5	.1	.3	.3	.2
Kurtosis	-.6	-.4	-.7	-.6	.4
Maximum	1.0000000	.9224092	.8205555	.8148572	.7173497
Minimum	.9539962	.8094354	.6617527	.6427498	.5473186
Actual Reliability = 1-MER	.991318	.951181	.903359	.898804	.849843
Distribution of R2MLE**					
Mean	.9905378	.9513628	.9042752	.8994287	.8489997
Standard deviation	.0048699	.0114074	.0147150	.0151360	.0164193
Skewness	-.9	-.2	-.3	-.3	-.2
Kurtosis	1.4	.0	-.2	-.2	-.0
Maximum	1.000000	.9823257	.9350639	.9329483	.8895143
Minimum	.9714401	.9100092	.8557628	.8498790	.7966162

*R1MLE is the estimate of 1-SCERD using proportional sampling.
**R2MLE is the estimate of 1-MER using random sampling.

the MLE of system reliability, using random sampling without replacement (R2MLE) and DUS sampling (R1MLE). However, two different reliability concepts were also used: R2MLE is an estimate of processing an account (or record) without a monetary error (i.e., complement of MER), while R1MLE is an estimate of processing a book value dollar without any compliance error (i.e., complement of SCERD).

An estimate of the proportion of times (out of 600 samples) the Mann 95% confidence lower bound on system reliability is less than actual system reliability for both random and DUS sampling is given in Table 11.

TABLE 11

ACTUAL CONFIDENCE FOR THE MANN NOMINAL 95% BOUND

Attribute Sampling Method	E1	E2	E3	E4	E5
DUS for attributes	.383333	.438333	.631666	.61	.551667
Random sampling for attributes	.605	.485	.493333	.483333	.513333

It is thus apparent that the theory for estimating system reliabilities from component data is deficient. However, the effect of this in terms of eventual impact on substantive tests is considerably mitigated by the linkage rules used in the simulation and the general audit setting--at least sufficiently so that the greater accuracy of the Mann method makes it more attractive than the overly conservative crude method. Thus the Mann method has potential for integrating the results of several tests concerning the internal control system. This is illustrated by the performances of the comprehensive audit strategies using this information and which are reported in subsequent subsections.

Comparison of stratified mean-per-unit testing with DUS testing

It appears that the best way to report on audit sampling strategy performance is to first consider the impact of the substantive testing method on that performance. The substantive test is the most important aspect of an audit strategy performance because, in the final analysis, it is the statistical validity of the test for the entire range of sample sizes available to the strategy that determines the validity of the audit strategy. Substantive test performance is thus the key factor behind the statistical validity of the internal control hypothesis.

As explained in chapter four, an attempt has been made to apply the substantive tests as is actually used in practice. One consequence of this approach is that the sample sizes for the two non-Bayesian substantive test methods differ considerably for the same degree of internal control reliance. This then makes the comparison of the impact of internal control information between strategies using different substantive test methods rather meaningless. For this reason this chapter is organized so that first the two non-Bayesian substantive methods are compared, and then the associated strategies using internal control information are reported in subsections dealing with particular substantive tests. Discussion of the potentially Bayesian and still theoretical Felix-Grimlund model seems more appropriate for separate treatment and so is put off for a separate subsection.

Chapter four indicated how the planned sample sizes are computed for the two non-Bayesian substantive testing methods, STMPU and DUS. Since these sample sizes vary depending upon the degree of reliance on internal controls, it appears that the best summary comparisons of these two methods can be made by considering their perofrmances for extremes of the sample sizes used by the audit strategies: this occurs with maximum reliance on internal controls (i.e., set $\beta = .5$) and with no reliance on internal controls (i.e., set $\beta = .05$).[5]

The planned sample sizes that result from these two β values using STMPU estimation are computed from the following formula given in p 217 for use when strata are divided into approximately equal amounts of book value:

$$n = \frac{z_\alpha^2 \, Y \sum_{i=1}^{20} N_i^2 \, \frac{\sigma_{Yi}^2}{Y_i}}{A^2 + z_\alpha^2 \sum_{i=1}^{20} N_i \, \sigma_{Y_i}^2}$$

(the variables are defined in chapter four). This formula requires knowledge of the book value, Y_i, for each stratum i and the associated variance of book values within the stratum, σ_{Yi}^2.

[5]Most sample sizes in the audit strategies tend to fall at these extremes anyway, except for linkage rule II(C) which results in a maximum $\beta = .2$ and hence a larger minimum sample size.

For the 20 strata used in the simulation of STMPU estimation, the these values are given in table 12.

TABLE 12

KEY STRATA CHARACTERISTICS

STRATUM NO. (i)	Ni	Yi	σ_{Yi}^2
1	7	656973.8	6493207.5
2	8	683348.2	7694542.4
3	9	655788.6	24574903.
4	12	710440.3	24896461.
5	15	695540.4	8499445.6
6	18	671065.8	8205089.2
7	22	690456.1	1773814.5
8	26	681708.2	1834924.6
9	32	702295.9	1166018.2
10	38	678094.7	1350068.9
11	48	684829.1	598006.68
12	61	689968.2	679513.77
13	78	684923.8	514699.43
14	102	680550.0	251643.04
15	137	684314.0	184296.88
16	189	684768.6	134672.65
17	278	683817.4	88765.289
18	444	684887.0	67173.457
19	871	685106.2	33444.521
20	4631	683626.7	19531.994

From table 12 it is apparent that the allocation procedure was successful in assigning approximately equal amounts of book value to each stratum.

Getting back to the sample size calculation formula for STMPU and using the values of table 12, $\Sigma N_i^2 \frac{\sigma_{Yi}}{Yi}^2 = 729007$ and $\Sigma N_i \sigma_{yi}^2 =$ 1,438,262,100. z_α in the formula is the standardized normal table value associated with $.5 - \alpha$ and is thus the complement of the α risk when using the positive approach. A is the planned precision based on controlling the α and the β risks at prespecified levels using the fromula $A = \frac{z_\alpha M}{z_\alpha + z_\beta}$ with the positive approach.[6] When using the positive approach, it should be noted that the only way internal control information affects the planned sample size is through the planned precision A.[7]

The calculation of the two extreme planned sample sizes using STMPU is as follows:

Assumed goals of simulated auditor

$\alpha = .05$, M = .05 x total book value = .05 x 13,671,503. = 683,575.

$\beta = .5$ for maximum reliance. Thus $A = \frac{z_{.05}}{z_{.05} + z_{.5}} (683,575) =$

$\frac{1.65}{1.65 + 0} (683,575) = 683,575$

and so

$$n = \frac{(1.65)^2 (13671503)(729007.97)}{(683575)^2 + (1.65)^2 (1438226100)} = 58$$

---

[6]These formulas reflect the positive approach. The equivalent negative apporach formula for sample size is:

$$n = \frac{z_\beta^2 Y \Sigma N_i^2 \frac{\sigma_{Yi}}{Yi}}{A^2 + z_\beta^2 \Sigma N i \sigma_{Yi}^2}$$

where now $A = \frac{z_\beta M}{z_\beta + z_\alpha}$

[7]The minimum planned sample size cannot be computed using the negative approach formula of footnote 6 because $z_{.5} = 0$ implying A=0,

For no reliance, $A = \frac{z_{.05}}{z_{.05} + z_{.05}}(683{,}575) = 1/2(683{,}575) = 341787.5$,

and so n = 225.

These sample sizes then need to be allocated to the strata in proportion to the book value of the strata. This results in almost equal samples from each strata. However, to assure statistical validity of the STMPU estimator, any fraction of a unit allocated to a stratum is automatically incremented to an additional unit.[8] This rule then results in the following extreme sample size from the STMPU estimator: maximum = 237 and minimum = 60.

Computation of sample sizes is much easier for DUS. Using the basic Poisson relationship of DUS, $B = nP$, the planned sample size is computed using the simple formula $n = \frac{B}{P}$ as explained on p. 220. Since the negative approach is conventionally used by DUS, the null hypothesis is that there is an exactly material amount of error. This means that the confidence level associated with the test is 1 - risk of Type II error = $1 - \beta$.[9] Thus for $\beta = .5$ the associated confidence level is .5 for DUS and for $\beta = .05$ the associated confidence level is

---

and so n becomes undefined at $\beta = .5$. However, the positive approach does yield a valid sample size of 58 for $\beta = .5$. For low values of β, both approaches yield the same planned sample size.

[8]This is to help assure that the actual sampling risks are held to their nominal level.

[9]The confidence level of a statistical test is the confidence associated with the test assuming the null hypothesis is true. Thus if the null hypothesis is that there are no errors (i.e., the positive approach) then the associated confidence level is $1 - \alpha$ for a one taile tailed test. On the other hand, if the null hypothesis is that there

.95 for DUS. The associated reliability factors, B, are B = 3.0 for confidence level of 95% and B = .6935 for confidence level of 50%.

As indicated in chapter four, the value for basic precision P is a function of materiality (.05), expected error, and precision gap widening factors. Using the actual planning rule of an auditing firm using the DUS approach, the planned basic precision values for each environment turned out to be the following:

Environment	E1	E2	E3	E4	E5
P value	.0448	.025	.025	.025	.025

It also turns out that the effect of using an increased basic precision for E1 is inappropriate for reasons discussed in the next subsection. Thus the same basic precision is used for all the environments when using DUS.[10]

The extremes for the planned sample sizes using DUS can now be computed. At maximum reliance on internal controls (50% confidence level) B = .6935, and $n = \frac{.6935}{.025} = 28$; and with no reliance, 95%

is an exactly material error (i.e., the negative approach) then the associated confidence level is 1 - β. Note that regardless of which approach is used, if the error condition is not exactly as assumed by the hypothesis, then actual confidence can be expected to be different from stated or nominal confidence.

[10]It is unclear how strongly this rule is held to in practice and how much to allow for precision gap widening factors. Certainly, the firm minimum precision = 1/2 of material dollar error rate works very well when compared to STMPU; and so this is the precision used in comparing DUS performance with STMPU. Note that always using a basic precision of 1/2 of materiality is the same as always using twice the discovery sample for the material error rate for whatever level of confidence is desired.

confidence level = B = 3.0, and so $n = \frac{3.0}{.025} = 120.$[11]

Upon seeing these sample sizes, comparing them to those of STMPU, and realizing they both result from a comparable set of goals, one may immediately begin to question the validity of such small sample sizes for DUS. The same doubts that Kaplan articulated about the "no free lunch hypothesis" come to mind and the suspicion is that DUS must be sacrificing something somewhere in order to result in such dramatically smaller planned sample sizes. That this is not the case is shown by the next four tables. Tables 13 and 14 show the actual sampling risks associated with DUS for the five environments.

The statistical performances of the STMPU estimator for its extreme sample sizes are given in table 15 and table 16.

To fully understand the use of these tables, it is perhaps best to begin by reviewing the statistics contained therein. Each table summarizes the results of a simulation of 600 substantive tests with the sample sizes indicated on each of the five environments. The key statistics as far as an audit strategy is concerned are the actual α and actual β risks and the average substantive test sample size (which in the four tables are constant). The actual α risk measures the proportion of times (out of 600) the test rejects the book value when there is an immaterial error (environments E1, E2, and E3), and the actual β risk measures the proportion of times the test accepts the

[11]Linkage II(C) results in maximum β = .2, hence the minimum sample size for this linkage is $n = \frac{R_{.8}}{.025} = \frac{1.6094}{.025} = 65.$

TABLE 13

DUS PERFORMANCE WITH MAXIMUM SAMPLE SIZE

Environment	Planned α	Actual α	Actual β	Planned β	1-Actual Confidence Level	Average Sample Size
E1: .1M	None Directly	.000	.000	.05	.000	120
E2: .5M	" "	.400	.000	.05	.000	120
E3: .99M	" "	.98999	.000	.05	.000	120
E4: 1.01M	" "	.000	.013333	.05	.000	120
E5: 1.5M	" "	.000	.000	.05	.000	120

TABLE 14

DUS PERFORMANCE WITH MINIMUM SAMPLE SIZE

Environment	Planned α	Actual α	Actual β	Planned β	1-Actual Confidence Level	Average Sample Size
E1:.1M	None Directly	.00333	.000	.5	.000	28
E2:.5M	" "	.40333	.000	.5	.000	28
E3:.99M	" "	.81999	.000	.5	.176667	28
E4:1.01M	" "	.000	.186667	.5	.188333	28
E5:1.5M	" "	.000	.05667	.5	.220	28

TABLE 15

STMPU PERFORMANCE WITH MAXIMUM SAMPLE SIZE

Environment	Planned α	Actual α	Actual β	Planned β	1-Actual Confidence Level	Average Sample Size
E1:.1M	.05	.021667	.000	.05	.046667	237
E2:.5M	.05	.451667	.000	.05	.051667	237
E3:.99M	.05	.95000	.000	.05	.041667	237
E4:1.01M	.05	.0000	.040	.05	.045000	237
E5:1.5M	.05	.0000	.000	.05	.038333	237

TABLE 16

STMPU PERFORMANCE WITH MINIMUM SAMPLE SIZE

Environment	Planned α	Actual α	Actual β	Planned β	1-Actual Confidence Level	Average Sample Size
E1:.1M	.05	.031667	.000	.5	.448333	60
E2:.5M	.05	.218333	.000	.5	.480	60
E3:.99M	.05	.500000	.000	.5	.488333	60
E4:1.01M	.05	.000	.476667	.5	.481667	60
E5:1.5M	.05	.000	.270000	.5	.505000	60

book value when there is a material amount of dollar error (environments E4 and E5). The environments are indicated on the left hand side of each table (the tables also indicate the amount of error in terms of proportion of materiality associated with each environment). The actual α and β risks are highlighted by the vertical lines separating them from the rest of the table. The planned α and β risks indicate the values used in deciding on the sample size and in constructing the statistical decision interval for the hypothesis test.[12]

The actual confidence level refers to the proportion of times the actual total audit value X is greater than the lower confidence limit $X \geq (\hat{X}-A')$ using the negative approach for STMPU, and the proportion of times $X \geq$ (Y-UB on E) using DUS.[13] This actual confidence level or "reliability" as Neter and Loebbecke called it represents one of the most important measures of statistical estimator used by earlier researchers. As can be seen by the actual α and β risks, however, the complement of this measure can be a poor indicator of these risks for different amounts of error.

The average sample sizes in tables 13, 14, 15, and 16 are exact because, in this case, constant sample sizes were used for each of the 600 simulations for each environment.

With this background it is now possible to discuss the information contained in tables 13, 14, 15, and 16.

---

[12]See chapter four for all formulas pertaining to the statistical test.

[13]See Neter-Loebbecke, pp. 110-11 for a demonstration of the equivalency of these two confidence intervals.

Perhaps the best starting point for discussion is to consider the apparently astonishing growth of the α risk for both substantive methods when the largest sample sizes are used. This perhaps is not as surprising for the DUS method which does not explicitly control for this risk as it is for the STMPU method which ostensibly does. The explanation lies in the kind of risk controlled for: in chapter four it is explained that the only Type I error controlled for at the nominal level with the present formulas used by auditors for STMPU is the error of rejecting the population when there are no errors in the population. As soon as there is some error the actual α risk can creep to beyond the nominal level. This is predicated by statistical theory in the manner the sample size formulas are coustructed for STMPU and DUS. Thus the results in the tables are perfectly in accord with statistical theory. However, since many auditors may not be aware of the assumptions underlying the use of sample planning formulas used in statistical decision making; appendix VII is provided to give an intuitive explanation as to why an auditor can expect such high α risks for certain levels of immaterial amounts of error, and why the maximum α risk is always the complement of the β risk (i.e., maximum α = 1 - maximum β).

With this understanding, what becomes surprising is not the high actual α level but the fact that for the maximum sample sizes, table 13 and table 15, the α risks are fairly comparable. This in spite of the fact that the maximum sample size for DUS is almost half the maximum sample size for STMPU. Thus Kaplan's no free lunch hypothesis appears to be rejected and the belief of DUS advocates confirmed.

Note that, statistically, both methods are fulfilling the predictions of their respective theories. For the STMPU method the α risk is held to within its nominal level when there are no or very few errors (as is the case in E1). However, as the error amounts increase within materiality the α risk climbs toward the limiting maximum which is 1 - actual β risk. This is the case for both the maximum and minimum sample sizes for the STMPU. Also note in tables 15 and 16 that the β risk is also held to within its nominal or planned level even for the highest such risk which occurs with the exactly material amount of error in E4.

Similarly, the performance of DUS is consistent with its theory. Note that maximum actual β risk is held to well within the planned β level. This is because as mentioned earlier (chpater two) the TACS evaluation procedure for DUS assures control of the β risk at the prespecified level even for the worst possible error pattern in terms of statistical detection for a given amount of total dollar error. Thus for most actual error patterns the maximum actual β will be less than the nominal level as is the case here. Again, note that the maximum α risk climbs toward its limiting level of 1 - actual β as explained in appendix VII.

Perhaps the most interesting and novel aspect of the relative performance of DUS with STMPU is the rate of climb of this actual α risk. Most arguments concerning these two substantive testing methods really revolve around this rate of climb of actual α risks and the point at which the two risks equal. It is evident from statistical theory that the α risk will be less for no errors with DUS than for STMPU, but greater for the maximum amount of immaterial error (E3) because $1 - \beta$ for DUS is greater than $1 - \beta$ for STMPU (since DUS minimizes β, it means that DUS also maximizes $1 - \beta$, the limiting value for α). Apparently, most classical sampling supporters feel that this crossover occurs with a relatively low proportion of immaterial errors (e.g., E1), whereas DUS advocates fell it occurs at a somewhat higher level (perhaps, for example, at E2). The simulation confirms the beliefs of the latter. In fact considering that at E2 the DUS α risk is less than the STMPU risk by an even greater amount than at #1, it appears that this crossover does not take place until the amount of immaterial error is close to materiality (say, over 2/3 materiality, to put it in an E3 type of environment). Thus, over most of the immaterial range the α risk is less for DUS when maximum sample sizes are used.

This is not the case for the minimum sample sizes, however, as a comparison between Tables 14 and 16 indicates. Here, the E2 α risk for STMPU is almost half the α risk for DUS. This appears to arise primarily because the range of α for STMPU has been cut in half by the $1 - \beta$ upper limit; whereas this is not the case for DUS. DUS does such a good job of keeping actual β down (nominal of .5 vs. actual of .186667) that the range of α is not as significantly reduced by the sample size reduction. Of course, it must also be remembered that the minimun DUS sample size is less than half that of the minimum

STMPU sample size, and for such a small sample size it is almost a wonder that α risks are not even greater. Moreover, it appears that for any sample size approximating that of STMPU, the DUS method will result in generally lower levels of α risk and assuredly lower levels of β risk. Nevertheless, the significant reduction of α risk which accompanies the reduction in sample size for STMPU estimation has important implications for the value of internal control information using this substantive testing method. This is further discussed in a later subsection.

To confirm the apparent superiority of DUS for the accounting environments used in the simulation, a simple experimental design was set up contrasting the two methods further. Both methods were compared using the maximum sample sizes for both methods. The maximum sample sizes were used because theoretically this is when the estimators should be performing the best.[14] The logic of the experimental design is illustrated in Table 17.

TABLE 17

BASIC EXPERIMENTAL DESIGN FOR
DUS VS. STMPU

Sample Size	DUS	STMPU
237	Looking for reduction in α risk (A)	Know α and β risks. Compare empirical with nominal. (B)
120	Looking for excessive α risk (C)	How badly does the estimator do in terms of empirical risks? (D)

[14]It is infeasible to use the minimum sample size of DUS with

Since parts C and B of the experimental design have already been reported in tables 13 and 15, respectively, tables 18 and 19 complete the evaluation of performance for the remaining two parts.

By comparing table 18 to table 15 and table 13 to table 19, it is evident that the DUS estimator continues to dramatically outperform STMPU. There thus appears to be little doubt that in many, if not all, auditing environments DUS is distinctly superior to STMPU for controlling the levels of <u>both</u> α and β risks. Teitlebaum has already proven analytically (and the simulation confirms) that the β risk is always less for DUS. What the simulation also shows is that this is substantially true for the α risk as well. Thus many of the prior fears expressed about incurring excessive α risk with DUS appear unfounded. When the sample sizes are equal DUS sampling risks are almost always less than STMPU, and even when sample sizes of DUS are almost 1/2 that of STMPU the risks may still be less (e.g., compare tables 13 and 15). Considering that the five accounting environments have characteristics similar to most of those in actual practice (see chapter two and chapter three), these results concerning the relative performance of DUS and STMPU may be the most important single finding of the present study. That is, for actual accounting systems DUS may be the best single statistical estimator for use in auditing, thus confirming the claims of DUS advocates all along.

STMPU because of the high degree of startification used here. That is, many strata would have only one observation thus making it impossible to construct an STMPU confidence interval. Hence a minimum sample size experimental design was not attempted.

TABLE 18

DUS PERFORMANCE WITH SAMPLE SIZE OF 237

Environment	Planned α	Actual α	Actual β	Planned β	1-Actual Confidence Level	Average Sample Size
E1	None Directly	.000	.000	.05	.000	237
E2	" "	.018333	.000	.05	.005	237
E3	" "	.986666	.000	.05	.006667	237
E4	" "	.000	.020	.05	.021667	237
E5	" "	.000	.000	.05	.051667	237

TABLE 19

STMPU PERFORMANCE WITH SAMPLE SIZE OF 120

Environment	Planned α	Actual α	Actual β	Planned β	1-Actual Confidence Level	Average Sample Size
E1	None Directly	.19500	.000	.05	.038333	120
E2	" "	.70333	.000	.05	.031667	120
E3	" "	.96333	.000	.05	.035	120
E4	" "	.000	.035	.05	.036667	120
E5	" "	.000	.001667	.05	.030	120

This discussion of the relative performance of DUS vs. STMPU is concluded with a set of statistics for the two estimators at the extreme sample sizes.

Table 20 provides the statistics for the DUS estimator. The table is organized as follows. First the estimated total errors from each environment are compared to the actual error amount at the top. Then the statistics on the upper bounds on error as described on p. 238 are provided. These bounds are 95% confidence bounds in the case of sample size of 120 (table 20), and 50% in the case of sample size of 28 (table 21). The statistics appear to support the conclusions reached on the performance of DUS in terms of α and β risks. For example, the fact that the mean of the estimated upper bounds on error is close to the actual amount considered material (683,575.15 = .05 x 13,671,503) for E2 predicts that about 40 to 50% of the time a Type I error will be made in E2--this is the α risk measured in the simulation. Also, as the number of errors found in the sample is increased (i.e., as one reads across the environments) the distribution of the upper bound on error as well as the MLE on error appears to tend toward normality (i.e., skewness and kurtosis tend toward zero, the estimator appears to be unbiased). As sample size is increased, the standard deviation is almost reduced in half for some of the estimates. This is to be expected by theory as the standard deviation of the mean estimates, $\sigma_{\bar{x}}$, is related to the standard deviation of the underlying variable, σ_x, as follows, $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$ when n is the sample size. Thus, if the sample size is quadrupled to 4n, the standard deviation of $\bar{x}$ is

TABLE 20

DUS SAMPLE SIZE 120 STATISTICS

	Environments				
	E1	E2	E3	E4	E5
Total error amount	37144.	318212.	672601.	688687.	1028369.
Distribution of Estimated total error					
Mean	34372.	323501.	678610.	693349.	1028004.
Standard deviation	46035.	91698.	195478.	195651.	280739.
Skewness	1.2	.7	.6	.4	.8
Kurtosis	.5	.3	-.3	-.5	-.3
Maximum	190148.	644978.	1279311.	1210400.	1801939.
Minimum	0.	132055.	333070.	313026.	552396.
Distribution of Upper bound on error					
Mean	376092.	672212.	1051259.	1069330.	1430095.
Standard deviation	46526.	97080.	211972.	210533.	313757.
Skewness	1.2	.7	.6	.4	.7
Kurtosis	.5	.3	-.4	-.5	-.4
Maximum	531450.	1014636.	1711890.	1625770.	2259597.
Minimum	341301.	473356.	674372.	654327.	893697.
Correlation between Estimated total error and Upper bound on error	.995	.986	.994	.994	.997

cut in half:

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{4n}} = 1/2(\frac{\sigma_x}{\sqrt{n}}).$$ (Compare table 20 to table 21.) The correlation can also be expected to be high since the two tend to move together with the amount of net error. These characteristics and the actual sampling risks tend to confirm the performance predicted by theory.

Tables 22 and 23 provide comparable statistics for the STMPU estimator at the extreme sample sizes of 237 and 60, respectively. Now the normality of the distribution of the estimator is more critical to predicting the performance of the estimator since the theory of STMPU and other classical estimators is based on the assumed normality of the estimator. In these two tables the indicated statistics are provided for the estimated total audit value and the upper confidence bound on the net error (95% confidence for sample size 237, table 22, and 50% confidence bound for sample size of 60, table 23). For the case of the upper bound on error, perhaps a more meaningful comparison is provided by looking at the actual total error for each environment indicated on table 21.

Overall, the distribution of the estimated total audit value appears to be approximately normal for both sample sizes, with the main difference between sample sizes being the expected 1/2 reduction in the standard deviation. This is also true for the distribution of the upper bound on error. It is interesting to note that the STMPU average upper bound on error is much smaller and hence more precise than that for DUS, but, unfortunately, it has much higher variability (standard

TABLE 21

DUS SAMPLE SIZE 28 STATISTICS

	Environments				
	E1	E2	E3	E4	E5
Total error amount	37144.	318212.	672601.	688687.	1028369.
Distribution of Estimated total error					
Mean	32031.	310198.	666467.	703962.	1016064.
Standard deviation	86921.	257711.	352945.	393415.	423611.
Skewness	2.6	.7	.2	.5	.3
Kurtosis	5.3	.1	-.3	.4	-.1
Maximum	507928.	1246171.	1700414.	2041200.	2304457.
Minimum	0.	0.	0.	-1145.	0.
Distribution of Upper bound on error					
Mean	370472.	649269.	1006798.	1044995.	1362306.
Standard deviation	86921.	258432.	353852.	391604.	422458.
Skewness	2.6	.7	.2	.5	.3
Kurtosis	5.3	.1	-.3	.4	.1
Maximum	846369.	1584613.	2030055.	2379641.	2832703.
Minimum	338442.	338442.	330442.	330442.	338442.
Correlation between Estimated total error and Upper bound on error	1.000	.999	.998	.998	.995

deviation) and so does not necessarily imply more reliable estimates.

At the bottom of the tables 22 and 23 are given the correlation between the estimated total audit value and the standard deviation of each sample. As discussed in chapter one, Kaplan has suggested that this may be an important measure for predicting the propensity of an estimator for making a Type II error. The researcher fails to see such a relationship, instead what appears to affect the correlation significantly is the monetary error rate, and perhaps the error pattern. The correlations presented here fall within the range charted by Neter-Loebbecke for STMPU for their different error patterns.

This then completes the description of the relative performance of the DUS and STMPU estimators. Generally, both methods perform as predicted even with the relatively small sample sizes associated with complete reliance on internal controls)which are about 1/4 the sample size used when there is no reliance--this is true for both estimators). Thus both estimators are sufficiently robust to allow for reliance on other information besides the substantive test. However, there is a significant difference in planned sample sizes and the Kaplan's free lunch hypotheses appears to be substantially correct for many auditing environments. That is, sample sizes using DUS can be significantly smaller for the same risk levels. Thus, for practical purposes, DUS appears to be a much more efficient substantive test method than STMPU.

The next subsections report on the performance of the audit sampling strategies using these estimators; and the effects of

TABLE 22

STMPU SAMPLE SIZE 237 STATISTICS

	Environments				
	E1	E2	E3	E4	E5
Total audit value	13634358.62	13353290.75	12998901.87	12982816.37	12643134.25
Distribution of estimated Total audit value					
Mean	13636356.	13364784.	13009455.	12994222.	12649634.
Standard deviation	211702.	217062.	232780.	232649.	238533.
Skewness	.3	.3	-.1	.1	.0
Kurtosis	-.2	-.3	-.1	.0	-.1
Maximum	14344292.	14006138.	13718670.	13766507.	13392552.
Minimum	13119141.	12831279.	12368127.	12304143.	11967507.
Distribution of Upper bound on error					
Mean	358739.	647050.	1025070.	1041203.	1403531.
Standard deviation	170842.	187677.	209571.	207693.	219213.
Skewness	-.4	-.3	.1	-.2	.0
Kurtosis	.1	-.3	-.0	.2	-.1
Maximum	744828.	1107855.	1563399.	1634835.	2039978.
Minimum	-265080.	82018.	358278.	309533.	707204.
Correlation between Estimated total audit value and Standard deviation	.726	.602	.512	.541	.458

TABLE 23

STMPU SAMPLE SIZE 60 STATISTICS

	Environments				
	E1	E2	E3	E4	E5
Total audit value	13634358.62	13353290.75	12998901.87	12982816.37	12643134.25
Distribution of estimated Total audit value					
Mean	13608822.	13341930.	13020241.	12956332.	12668217.
Standard deviation	402009.	417471.	461598.	460594.	504258.
Skewness	.4	.3	.4	.2	.2
Kurtosis	.0	-.3	.1	.0	-.3
Maximum	14974969.	14540167.	14585704.	14475337.	14141690.
Minimum	12709789.	12332462.	11843032.	11646991.	11264122.
Distribution of Upper bound on error					
Mean	62680.	329573.	651261.	715170.	1003285.
Standard deviation	402009.	417471.	461598.	460594.	504258.
Skewness	-.4	-.3	-.4	-.2	-.2
Kurtosis	.0	-.3	.1	.0	-.3
Maximum	961714.	1339040.	1028471.	2024512.	2407381.
Minimum	-1303466.	-876664.	-914201.	-803834.	-470187.
Correlation between Estimated total audit value and Standard deviation	.610	.491	.412	.354	.285

internal control information and linkage rules on substantive test sample size, actual α, and combined risks. First, all audit sampling strategies using DUS for substantive testing are reported.

Performance of audit strategies
using DUS

This subsection reports on all strategies using DUS for the substantive testing stage of a strategy. As discussed in the preceding subsection, this is the best performing of the substantive testing methods in the dissertation, and hence the smallest average substantive test sample sizes result from strategies using this method.

The preceding subsection has also reported the performance of the "informationless" strategy using DUS with no internal control information. That is, in effect table 13 presents the performance of Strategy 14 as described on p.230 because without internal control information $B = 3:0$; $P = .025$ as discussed on pp. 267 , and so $n = \frac{3.0}{.025} = 120$. This is the constant sample size used by Strategy 14 on all five environments.

To show what is the effect of having perfect information about internal controls (i.e., omniscience in the sense discussed on p 223-5), tables 24, 25, and 26 present the performances of audit strategies with this information and using linkage rules II(A), II(B), and II(C), respectively. In addition table 13 is repeated to facilitate the contrast of the effects of the internal control information.

Again, perhaps it is best to first review the various parts of the tables. The format is very similar to that of the preceding

TABLE 13

PERFORMANCE OF STRATEGY 14: E - B

(Informationless, no linkage with DUS)

Environment	Planned α	Actual α	Actual β	Planned β	1-Actual Confidence Level	Average Sample Size
E1	None directly	.000	.000	.05	.000	120
E2	None directly	.400	.000	.05	.000	120
E3	None directly	.989999	.000	.05	.000	120
E4	None directly	.000	.013333	.05	.000	120
E5	None directly	.000	.000	.05	.000	120
Realistic Average		.2685	.000333			120
Conservative Average		.278	.002666			120

TABLE 24

PERFORMANCE OF STRATEGY 3: AAB
(OMNISCIENCE, ELLIOTT AND ROGERS LINKAGE WITH DUS)

Environment	Planned α	Actual α	Actual β	Planned β	1-Actual Confidence Level	Average Sample Size
E1	None Directly	.003333	.000	.5	.000	28
E2	" "	.403333	.000	.5	.000	28
E3	" "	.986666	.000	.1	.011667	93
E4	" "	.000	.013333	.05	.000	120
E5	" "	.000	.000	.05	.000	120
Realistic Average		.270666	.000333			42.35
Conservative Average		.278666	.002666			77.8

TABLE 25

PERFORMANCE OF STRATEGY 4: ABB
(OMNISCIENCE, SAS NO. 1 LINKAGE WITH DUS)

Environment	Planned α	Actual α	Actual β	Planned β	1-Actual Confidence Level	Average Sample Size
E1	None Directly	.003333	.000	.5	.000	28
E2	" "	.403333	.000	.5	.000	28
E3	" "	.986666	.000	.5	.176667	28
E4	" "	.000	.013333	.05	.000	120
E5	" "	.000	.000	.05	.000	120
Realistic Average		.270666	.000333			32.6
Conservative Average		.278666	.002666			64.8

TABLE 26

PERFORMANCE OF STRATEGY 5: ACB
(OMNISCIENCE, CLARKSON, GORDON & CO. LINKAGE WITH DUS)

Environment	Planned α	Actual α	Actual β	Planned β	1-Actual Confidence Level	Average Sample Size
E1	None Directly	.000	.000	.2	.000	65
E2	" "	.363333	.000	.2	.000	65
E3	" "	.886666	.000	.2	.000	65
E4	" "	.000	.013333	.05	.000	120
E5	" "	.000	.000	.05	.000	120
Realistic Average		.2419998	.000333			67.75
Conservative Average		.249999	.0026666			87

subsection. The main difference now is that substantive test sample sizes between environments are allowed to vary as determined by the linkage rule but within an environment they are constant. Due to the way omniscience has been defined in this dissertation (i.e., as an upper bound on greater and greater accuracy in obtaining reliability estimates using a particular linkage rule), tables 24, 25, and 26 can be largely constructed from tables 13 and 14 which report on many of the sample sizes used by the strategy.

Note that although it is not indicated in these tables, the planned combined risk is .05 for all tables. The actual combined risk for a strategy is the same as the actual β risk column because these β risks are the result of the planned β risk associated with an environment; which in turn are dependent on the interaction of amount of internal control information, linkage rule, and the planned combined risk. When the first two are held constant, the actual β risk also measures the actual combined risk. Thus for example, in table 25 the planned β risk for E2 is .5, planned combined risk is .05, and actual β and combined risk is 0 because there is an immaterial amount of error in E2.

The two averages for actual α and β risks and the average substantive test sample size on the bottom reflect the weighting schemes discussed on pp.233-7. They are useful for illustrating the sensitivity of summary measures of performance to different assumptions about the frequencies with which these five environments occur in actual practice. They can be interpreted as long run average actual risks and sample sizes associated with a strategy, and thus provide a summary measure over all environments as discussed in section 5.2.

The major conclusions that can be reached upon contrasting the four tables appear to be the following: First, by contrasting tables 24, 25, and 26 with table 13; it is clearly evident that the weak and strong forms of the internal control hypothesis of auditing are upheld for all non-Bayesian linkage rules using DUS. Second, the most efficient linkage in terms of maximum reduction of sample size and maintenance of the sampling risks is Linkage II(B)--the SAS No. 1 linkage. However, in general, this performance may be misleading because with imperfect information this linkage can lead to excessive combined risk levels (i.e., there is no inherent conservatism to put constraints on the frequency of unwarranted reliance). This problem is compounded by the fact that the approximation of lower bound on system reliability provided by the Mann method on which the C value is based, is biased towards optimism, i.e., higher bounds are computed than indicated by the sampling distribution (see section 5.3). Nevertheless, with better and more accurate information, linkage II(C) represents the maximum impact internal control information can make on substantive tests. This means that with good internal control information it is statistically valid to reduce the sample size as much as 75% when using DUS.

Thus the rankings of linkages in terms of impact on sample size is first II(B), then II(A), and, finally, II(C). Thus, in a statistical sense, the firm of Clarkeson, Gordon, & Co. can potentially increase the value of internal control information by adopting a less conservative linkage rule. However, this is conditional on the fact that a certain degree of quality about the internal control information is assured.

The use of the summary averages does not provide any additional insights than that already given in terms of linkage rankings. Generally, the average α and β risks for linkage rules are comparable. Only the sample sizes show significant differences among linkages, and again those associated with the SAS No. 1 linkage being the smallest (up to almost 1/4 the "informationless" sample size of Strategy 14).

Tables 27, 28, and 29 provide more realistic performances of the linkages. Now, the internal control information is imperfect being based on sample evidence from compliance testing and only an approximation of the lower bound on system reliability. Thus these tables give the results of a comprehensive simulation of an audit strategy starting with compliance testing, linkage of the compliance test results to the amount of substantive testing, and a statistical decision based on the resulting substantive test. Thus the compliance test results now determine audit strategy performance. Tables 27, 28, and 29 are best compared to tables 24, 25, and 26, respectively. The results of these comparisons indicate, as expected, an increase in the average substantive test sample size because of increased false alarm risk (but note not nearly to the extent indicated by the discussion on pp166-9 which deals primarily with the statistical theory associated with a single compliance test) as well as by increased β or combined

TABLE 27

PERFORMANCE OF STRATEGY 9: BAB
(OBJECTIVE, ELLIOTT AND ROGERS' LINKAGE WITH DUS)

Environment	Planned α	Actual α	Actual β	Planned β	1-Actual Confidence Level	Average Sample Size
E1	None Directly	.003333	.000	.5	.000	28
E2	" "	.391667	.000	.37	.000	39.11
E3	" "	.9800	.000	.09	.016667	97.90167
E4	" "	.000	.031667	.08	.033333	102.83
E5	" "	.000	.000	.05	.031667	120.00
Realistic Average		.266166	.000792			45.99
Conservative Average		.275	.006333			77.57

TABLE 28

PERFORMANCE OF STRATEGY 10: BBB
(OBJECTIVE, SAS NO. 1 LINKAGE WITH DUS)

Environment	Planned α	Actual α	Actual β	Planned β	1-Actual Confidence Level	Average Sample Size
E1	None Directly	.0033333	.000	.5	.000	28
E2	" "	.406667	.000	.5	.000	28.97
E3	" "	.956666	.000	.08	.035	105.74
E4	" "	.000	.025	.06	.026667	111.69
E5	" "	.000	.000	.5	.031667	120.
Realistic Average		.267167	.0006			44.34
Conservative Average		.273331	.005			78.88

TABLE 29

PERFORMANCE OF STRATEGY 11: CCB
(OBJECTIVE, CLARKSON, GORDON, & CO. LINKAGE WITH DUS)

Environment	Planned α	Actual α	Actual β	Planned β	1-Actual Confidence Level	Average Sample Size
E1	None Directly	.000	.000	.2	.000	65
E2	" "	.373333	.000	.15	.000	78.66
E3	" "	.989999	.000	.05	.000	120
E4	" "	.000	.013333	.05	.015	120
E5	" "	.000	.000	.05	.031667	120
Realistic Average		.2604	.000333			86.10
Conservative Average		.272666	.002666			100.73

risk as a result of unwarranted reliance due to sampling error. Overall, though, the results are very similar to the omniscient case of tables 24, 25, and 26.

There are perhaps two noteworthy aspects of the objective case strategies with DUS. One is that the superiority of the SAS No. 1 has evaporated under uncertain information and the conservatism of the linkage now actually boosts the performance of the strategy (this will become particularly evident when discussing STMPU estimation). Second, under uncertainty the conservatism of the Clarkeson, Gordon, & Co. linkage has become even more evident as the sample size for E3 remains a constant at 120. This really reflects the additional conservatism of the three times materiality rule since table 26 reflects only the conservatism due to not allowing planned β risk to climb higher than .2. Thus there are two components to the DUS linkage rule: (1) β risk not allowed to go higher than .2 which indicates an approximate additional substantive testing of about 25-35 sample items (using the realistic weighting); and (2) the three times materiality rule which induces an additional testing of about 20 sampling units (difference between realistic average sample size of tables 29 and 26). However, again, in the case where there is sufficient uncertainty associated with internal control evaluation, a conservative linkage could prove to be optimal. This will be made evident soon.

To allow a better assessment of linkage performance, some additional statistics were computed for the sample size distributions of the various linkage rules. This was done for all environments for which the sample sizes were not constant for all simulated internal control evaluations. These statistics are listed by linkage rule (indicated by table and strategy number) and environment. In all tables NSMPL stands for the sample size.

Generally, the tables do not provide evidence that the rankings of the linkage rules should be reconsidered. It turns out that the best performing linkage appears to be the Elliott and Rogers Rule. Although when used with DUS the SAS No. 1 linkage compares well with the Elliott and Rogers linkage, this is not the case when STMPU is the substantive testing method (this is shown in the next subsection). Thus, the more conventional Elliott and Rogers linkage is used to simulate the impact of judgmental errors in interpreting internal control information on the performance of an audit strategy. This performance is summarized in table 34. The compliance test results are exactly the same as in table 28 (strategy 10) except now a random disturbance factor is added to the estimated R value before the linkage rule is applied (see pp.174-6 for more explanation of this strategy). Note that this disturbance factor results in simulating a suitably confused auditor since the average sample size for E4 is less than that for E3--this indicates an auditor uncertain about the environment facing him.

Comparing table 34 with table 28, its objective case counterpart, several differences stand out. First note the very large growth in the actual β risk (combined risk) to well beyond even the nominal level. The increase in risk is about 10 times that of table 28. This

TABLE 30

STRATEGY 9: TABLE 27 SAMPLE SIZE STATISTICS

	Environments		
	E2	E3	E4
Distribution of NSMPL*			
Mean	39.1133	97.9017	102.8300
Standard deviation	12.4876	20.4945	18.8837
Skewness	.82	-.18	-.38
Kurtosis	.52	-1.25	-1.52
Maximum	93.0000	120.0000	120.0000
Minimum	28.0000	49.0000	49.0000

*NSMPL is used to represent the variable, sample size.

TABLE 31

STRATEGY 10: TABLE 28 SAMPLE SIZE STATISTICS

	Environments		
	E2	E3	E4
Distribution of NSMPL			
Mean	28.9717	105.7417	111.6850
Standard deviation	9.2434	31.9652	25.2145
Skewness	9.65	-1.87	-2.80
Kurtosis	92.33	1.66	6.13
Maximum	120.0000	120.0000	120.0000
Minimum	28.0000	28.0000	28.0000

TABLE 32

STRATEGY 11: TABLE 29 SAMPLE SIZE STATISTICS

	Environments E2
Distribution of NSMPL	
Mean	78.6583
Standard deviation	23.7824
Skewness	1.2
Kurtosis	-.64
Maximum	120.0000
Minimum	65.0000

TABLE 33

STRATEGY 16: TABLE 34: SAMPLE SIZE STATISTICS

	Environments				
	E1	E2	E3	E4	E5
Distribution of NSMPL					
Mean	64.3733	69.6400	79.4817	77.7083	84.1850
Standard deviation	43.9140	44.3049	44.5362	44.2636	43.6620
Skewness	.4	.2	-.2	-.2	-.4
Kurtosis	-1.77	-1.91	-1.90	-1.92	-1.75
Maximum	120.0000	120.0000	120.0000	120.0000	120.0000
Minimum	28.0000	28.0000	28.0000	28.0000	28.0000

TABLE 34

PERFORMANCE OF STRATEGY 16: DAB
(SUBJECTIVE, ELLIOTT AND ROGERS LINKAGE WITH DUS)

Environment	Planned α	Actual α	Actual β	Planned β	1-Actual Confidence Level	Average Sample Size
E1	None directly	.005000	.000	.2	.000	64.37
E2	" "	.401666	.000	.19	.000	69.64
E3	" "	.915000	.000	.14	.0783333	79.48
E4	" "	.000	.101667	.15	.101667	77.71
E5	" "	.000	.021667	.13	.1000000	84.19
Realistic Average		.26025	.003083			69.05
Conservative Average		.264333	.024667			75.08

increased risk is accompanied by average sample sizes either comparable to that of a knowledgeable auditor (conservative average) or even greater than those used by a knowledgeable auditor (realistic average). Thus a nonknowledgeable or an insufficiently trained auditor who attempts to use internal control information to reduce substantive testing is not only likely to increase his overall combined risk, but this could occur with an overall increase in substantive testing compared to a knowledgeable auditor. Thus with sufficient judgmental errors the internal control hypothesis is violated.

Note, however, that the α risk is hardly affected by the judgmental errors.

The conclusions that these results appear to point to are the following. Although the statistical validity of the internal control hypothesis has been confirmed, in actual practice the hypothesis may be violated. Certainly, a conservative linkage such as the Clarkeson, Gordon & Co. one, which significantly reduces the maximum reliance allowed on internal controls (thus implicitly recognizing the increased uncertainty associated with this information), and hence limits substantive sample size reduction; would have a lower β risk in this situation. In fact for certain levels of subjective errors it may thus prove to be an optimal linkage.

An aspect of DUS which has been touched upon earlier and which is now analyzed in more detail is the apparently frequent practice of increasing the basic precision when the expected error is a small proportion of materiality. This is another manifestation of the auditing philosophy that a clean audit ought to require less testing than a dirty audit. In this case the degree of dirtiness is measured by the monetary error rate. The end result of this practice is to reduce planned sample size (since $n = \frac{R}{P}$, if P = basic precision is allowed to increase, then n becomes smaller) for the statistical tests. Again, the implied assumption behind this practice is that the sample size can be decreased without any appreciable increase in audit risks.

However, the results of the simulation do not bear out this assumption as is indicated by the α risk for E1 when basic precision is increased as a result of recognizing the exceptionally low monetary error rate in this environment (.1 of materiality). This is done by adjusting (subtracting) from materiality (.05) the expected dollar error rate (in E1 this is .0027169) and any precision gap widening factors. (See pp 219-20 for a discussion of this procedure and references for it). To be conservative let the adjustment be equal to the error rate .0027169 (according to Clarkeson, Gordon & Co. this adjustment is usually only 1/2 of the most likely errors). Then basic precision P = .05 - 2(.0027169) = .044566 is a conservative basic precision (by letting the sample size allow for more than the expected number of errors in the sample).

This basic precision is now considered for three different situations: one where reliance on internal controls is maximum, B = .6935, one where the maximum reliance is at a higher confidence level, B = 1.6094, and one where there is no reliance on internal controls, so B = 3.0. This results in three sample sizes for this basic precision adjustment performance:

$n_1 = \frac{.6935}{.044566} = 16$, $n_2 = \frac{1.6094}{.044566} = 36$, $n_3 = \frac{3.0}{.044566} = 67$. The performance of these sample sizes is now compared with that of the sample sizes used without the adjustment to basic precision. This results in sample sizes of 120 for B = 3.0 (see table 13), 65 for B = 1.6094 (see table 26), and 28 for B = .6935 (see table 25). Now the actual α risks associated with these sample sizes and related confidence levels is given in table 35.

TABLE 35

EFFECT OF ADJUSTING BASIC PRECISION USING DUS IN E1

	Confidence Levels		
	95% (B=3.0)	80% (B=1.6094)	50% (B=.6935)
Without Adjustment	n=120, α=000000	n=65, α=000000	n=28, α=.003333
With Adjustment	n=67 α=.265000	n=36, α=.228333	n=16 α=.098333

It is thus clearly evident that adjusting basic precision to above the smallest amount can be an inappropriate technique of DUS. More generally, table 35 appears to indicate that any reduction of sample size without an associated reduction in the confidence level (where confidence level = 1 - β under the negative approach and this is soon shown to be the only appropriate approach in auditing) will automatically and significantly increase the α risks associated with the test.

This phenomenon became apparent to the researcher early in the simulation when it became evident that unless sample size and confidence level were properly matched, the statistical decision rule would not perform as predicted. This is especially important when sample size is reduced as a result of internal control information. Thus if the sample size reduction is not accompanied by the requisite confidence level reduction (as determined by statistical theory via sample computation formulas), the α risk can increase significantly. (For an example of this phenomenon using STMPU, compare the actual α risks of table 15 with that of table 19). However, an increase in sample size does not appear to require an accompanying increase in the confidence level (compare table 9 with table 14). The risks only seem to decrease in such a situation.

Finally, mention should be made of the relationship of the actual confidence level to the sampling risks. Generally, under the negative approach the confidence level $(1 - \beta)$ may give an indication of the risk of Type II error at the material level. However, it is not a good indicator of how the risk varies with differing amount of errors (for example, see table 14, E5 performance) and it is no indicator of the α risk associated with an audit strategy. Thus it is not surprising that so little could be concluded about relative strategy performances just on the basis of the actual confidence level. Interpretation problems become even worse if the positive approach is used in setting up the statistical decision. This is further discussed in the next subsection.

Performance of audit strategies
using STMPU

This subsection reports on all strategies using STMPU as the substantive testing method. Although the sample sizes using this method tend to be about twice as large as that for DUS, the overall reduction in substantive test sample size as a result of internal control reliance works out to be the same proportion as in DUS (i.e., maximum reliance results in about a 75% sample size reduction just as in DUS). However, unlike DUS there are significant differences in sampling risks which accompany the reduction in sample size.

A previous subsection (p263-289) reports on the comparison of DUS with STMPU performance. There the sample sizes with minimum and maximum reliance on internal controls are computed and found to be 237 and 60, respectively. Table 15 gives the performance of the "informationless" strategy not using internal control information so that there is no reliance and, hence, the maximum sample size applies to all environments. This is in effect strategy 13: E - A.

To show what is the effect of having perfect information about internal controls (i.e., omniscience in the sense discussed on pp.154-6), tables 36 and 37 present the performance of audit strategies using this information and using linkage rules II(A) and II(B) respectively. In addition, for convenience, table 15 is repeated to facilitate the comparison of the effects of internal control information.

The meanings of the terms and statistics in the tables have been described in the preceding subsections.

These tables show that, again, the statistical validity of both

TABLE 36

PERFORMANCE OF STRATEGY 1: AAA

(Omniscience, Elliott and
Rogers Linkage with STMPU)

Environment	Planned α	Actual α	Actual β	Planned β	1-Actual Confidence Level	Average Sample Size
E1	.05	.031667	.000	.5	.448333	60
E2	.05	.218333	.000	.5	.48000	60
E3	.05	.906667	.000	.1	.088333	183
E4	.05	.000	.040	.05	.045	237
E5	.05	.000	.000	.05	.038333	237
Realistic Average		.217333	.001			87.3
Conservative Average		.231333	.008			155.4

TABLE 37

PERFORMANCE OF STRATEGY 2: ABA
(Omniscience, SAS No. 1 Linkage with STMPU)

Environment	Planned α	Actual α	Actual β	Planned β	1-Actual Confidence Level	Average Sample Size
E1	.05	.031667	.000	.5	.448333	60
E2	.05	.218333	.000	.5	.480	60
E3	.05	.500	.000	.5	.488333	60
E4	.05	.000	.040	.05	.045	237
E5	.05	.000	.000	.05	.038333	237
Realistic Average		.156334	.001			68.85
Conservative Average		.2905	.008			130.8

TABLE 15

PERFORMANCE OF STRATEGY 13: E - A

(Informationless, No Linkage with STMPU)

Environment	Planned α	Actual α	Actual β	Planned β	1-Actual Confidence Level	Average Sample Size
E1	.05	.021667	.000	.05	.046667	237
E2	.05	.451667	.000	.05	.051667	237
E3	.05	.950	.000	.05	.041667	237
E4	.05	.000	.040	.05	.045000	237
E5	.05	.000	.000	.05	.038333	237
Realistic Average		.288833	.001			237
Conservative Average		.284667	.008			237

the strong and weak forms of the internal control hypothesis are being supported. Sample size can be reduced as much as 75% without increasing sampling risks. In fact, the α risk is dramatically cut in half when internal controls can be relied upon. The reason for this is simple and is based on statistical theory as outlined on appendix VII. Since the upper bound on the α risk must be 1 - maximum β and with maximum reliance β can go as high as .5, this upper bound becomes .5 instead of .95 with no reliance. Thus the range of the β risk is cut in half and so, accordingly, are the risks over the range of immateriality (compare table 25 to table 11).

But then the question might arise why this did not also occur with DUS. The answer is that it does, but that because DUS controls the actual β risk so well within the nominal level (particularly when the planned β is high); the complement, 1 - β, and, hence, α is automatically higher. Thus there is always a certain amount of trade off between α and β risk that occurs in evaluating an audit strategy. Of course, it must also be remembered that the minimum sample size for DUS is less than half that for STMPU.

Nevertheless, the implications of the finding that there is a significant reduction in α risk as well as in sample size with internal control reliance, implies that the value of internal control information with STMPU may be much higher than when using DUS. This might imply that firms using STMPU would rely on internal controls more frequently and to a greater extent than firms using DUS.

On the other hand, the fact that the actual α is closer to the nominal level also means that the risk of unwarranted reliance (which occurs with imperfect information about the internal controls) has a much greater chance of forcing the combined risk to go above the nominal level of .05. Hence, if the linkage rules themselves are not sufficiently conservative to counteract the unwarranted reliance effect effect, actual combined risk may go beyond the nominal level. Since the simulation attempted to limit the conservatism of linkages II(A) and II(B) to obtain a higher upper bound on the impact of internal control information, the unwarranted reliance effect was sufficient to perceptibly increase the proportion of Type II errors to greater than the nominal .05 level. This is shown in tables 38 and 39 which are based on objective, imperfect information about internal controls from compliance testing and using the approximate Mann bound.

The insufficient conservatism in both the linkage rule and the substantive test thus allows the β risk to be higher than nominal. Although in a strict sense using the definition of significant difference given in pp.237 one might argue both strategies 7 and 8 uphold the internal control hypothesis by not being more than 2.5 percent points beyond the planned level, it is clear that the risk of Type II error has increased. Such increase risk can be planned for and limited by making the linkage rules more conservative (as indicated on p.181) or by directly using a lower planned β level for the substantive test as suggested by Don Roberts. Note, however, that such adjustments are not necessary for DUS because it is sufficiently conservative in controlling for actual β risk that even

TABLE 38

STRATEGY 7: BAA

(Objective, Elliott and Rogers Linkage with STMPU)

Environment	Planned α	Actual α	Actual β	(Average) Planned β	1-Actual Confidence Level	Average Sample Size
E1	.05	.0316667	.000	.5	.448333	60
E2	.05	.253333	.000	.37	.343333	82.79
E3	.05	.888333	.000	.09	.110	196.49
E4	.05	.000	.068333	.07	.073333	205.35
E5	.05	.000	.001667	.05	.038333	237.00
Realistic Average		.225083	.00175			95.37
Conservative Average		.234666	.014			156.33

TABLE 39

STRATEGY 8: BBA

(Objective, SAS No. 1 Linkage with STMPU)

Environment	Planned α	Actual α	Actual β	(Average) Planned β	1-Actual Confidence Level	Average Sample Size
E1	.05	.031667	.000	.5	.448333	60
E2	.05	.225000	.000	.48	.463333	61.87
E3	.05	.906667	.000	.07	.090000	209.92
E4	.05	.000	.076667	.06	.076667	221.22
E5	.05	.000	.000	.05	.036667	237.00
Realistic Average		.219334	.061917			91.505
Conservative Average		.232667	.015333			158.00

certain increases in the risk of unwarranted reliance are not sufficient to increase combined risk above the nominal .05 level. Again, though, strategy performance is dependent on the degree of uncertainty concerning reliance on internal controls, for with enough uncertainty even DUS strategies fail to conform to the hypothesis (compare table 34 to table 27).

At the bottom of each page are the summary averages associated with each strategy. Again, they may be interpreted as a rough estimate of actual α and combined risks that are associated with applying a particular strategy over the long run. Generally, these are consistent with the rest of the table and highlight the fact that sampling error in compliance testing tends to make the Elliott and Rogers Linkage more attractive because it inhibits β and combined risk growth due to unwarranted reliance, which automatically rises with imperfect information about internal controls. This is somewhat different from results obtained with DUS which did not show much change in the performances of the two linkages between the omniscient and objective internal control information alternatives. Here the relationships are reversed (compare summary averages of tables 36 and 37 to tables 38 and 39, respectively). These results confirm the choice of the Elliott and Rogers linkage as the best of the objective case linkages.

A review of the relationship of the 1 - (actual confidence levels) with α and β risks again shows more relationship with the β risk at the exact materiality level E4. However, this measure provides little guidance for measuring the actual α risk level for different levels of immateriality.

Perhaps it is most appropriate at this time to consider the issue of the positive vs. the negative approach to hypothesis testing when using STMPU (the negative approach is always used with DUS). As explained on pp. 214 of chapter four and appendix VI, the confidence level associated with a statistical test is a function of the approach used in setting up the null hypothesis. Generally, under the negative approach the nominal confidence level = 1 - planned β level; whereas, under the positive approach the nominal confidence level = 1 - planned α level where α and β have been defined on p. 55 of chapter two.

It turns out that the only statistically valid approach for auditor decision making is the negative approach since the positive approach approach as described on pp.214-16 of chapter four is identical to the negative approach. This is because the adjustment to achieved precision described on p217 in effect amounts to converting the positive approach to a negative approach in terms of converting the confidence level associated with the hypothesis test from one equal to $1 - \alpha$, to $1 - \beta$. The effect of this adjustment to precision on strategy performance is illustrated in tables 40, 41, 42, and 43. Tables 40 and 42 result from using the positive approach described on pp. 217 . By comparing these tables with tables 15 and 16, it is evident that this positive approach is equivalent to the negative approach. This is because the use of the adjusted precision rather than the achieved precision to control the β risk at the planned level is tantamount to changing the approach. If no such adjustment is made and the

TABLE 40

STMPU USING THE POSITIVE APPROACH
WITH ADJUSTED PRECISION A''

Environment	Planned α	Actual α	Actual β	Planned β	1-Actual Confidence Level	Average Sample Size
E1	.05	.021667	.000	.05	.046667	237
E2	.05	.451666	.000	.05	.051667	237
E3	.05	.950	.000	.05	.041667	237
E4	.05	.000	.040	.05	.045	237
E5	.05	.000	.000	.05	.038333	237

TABLE 41

STMPU USING THE POSITIVE APPROACH
WITH ACHIEVED PRECISION A'

Environment	Planned α	Actual α	Actual β	Planned β	1-Actual Confidence Level	Average Sample Size
E1	.05	.141667	.000	.05	.046667	237
E2	.05	.470	.000	.05	.051667	237
E3	.05	.858333	.000	.05	.041667	237
E4	.05	.000	.118333	.05	.045	237
E5	.05	.000	.006667	.05	.038333	237

TABLE 42

STMPU USING THE POSITIVE APPROACH
WITH ADJUSTED PRECISION A"

Environment	Planned α	Actual α	Actual β	Planned β	1-Actual Confidence Level	Average Sample Size
E1	.05	.031667	.000	.5	.043333	60
E2	.05	.218333	.000	.5	.035000	60
E3	.05	.500	.000	.5	.058333	60
E4	.05	.000	.476667	.5	.043333	60
E5	.05	.000	.270	.5	.061667	60

TABLE 43

STMPU USING THE POSITIVE APPROACH
WITH ACHIEVED PRECISION A'

Environment	Planned α	Actual α	Actual β	Planned β	1-Actual Confidence Level	Average Sample Size
E1	.05	.210	.000	.5	.043333	60
E2	.05	.320	.000	.5	.035000	60
E3	.05	.460	.000	.5	.058333	60
E4	.05	.000	.510	.5	.043333	60
E5	.05	.000	.371667	.5	.061667	60

hypothesis test is based on the achieved precision (this will be called here the pure positive approach), the STMPU performance becomes as indicated in tables 41 and 43.

Note that under the pure positive approach both actual α and β risks can go to well beyond the nominal levels predicted by the theory. Thus the internal control hypothesis is not statistically valid with the pure positive approach and the negative approach is the only valid approach in auditing.

Also note that this result is consistent with the general rule developed earlier in the chapter. That is, sample size cannot be reduced unless the confidence level (which is reflected by the associated confidence coefficient used in the statistical hypothesis test: B in DUS and Z in STMPU) is also reduced, or equivalent (in the case of STMPU under positive approach, the equivalent is the adjustment for achieved precision). Thus the rule appears to apply to all sampling procedures.

An interesting aspect of this general rule is that the confidence level associated with a statistical test is not necessarily an informative indicator of substantive test method performance unless the correct confidence measure is used. In particular it appears that the confidence level under the positive approach is not very useful for predicting either the actual α risk level with various amounts of immaterial errors or the actual β risk level for any degree of material errors. To appreciate this consider the complements of the actual confidence levels in tables 40 and 42 and compare them to actual α and

β risks in both tables--note that the actual confidence levels hardly vary for all of these situations. Since most earlier studies have used this actual confidence measure under the positive approach (at least for the classical estimators) to compare performances it is not surprising that they failed to provide insights on the issues addressed here.

This subsection concludes with a presentation of the characteristics of the sample sizes used in the objective strategies in tables 44 and 45. Again, these statistics do not appear to lead to a reconsideration of linkage rule performance.

Performance of audit strategies
using the Felix-Grimlund model

The Felix-Grimlund substantive testing method is at present mainly of theoretical interest since it is not being used in practice and its performance under various error conditions has not yet been empirically assessed. As described in chapter four, it appears that the most feasible approach using this model is to adapt it to DUS. This is what was done in the simulation. Thus the most relevant comparison of performance is to the traditional DUS approach.

Since the maximum planned sample size using DUS is 120 units, it was decided to use this as a ceiling in planning sample sizes using the Felix-Grimlund model. Otherwise, for large enough total dollar error the Felix-Grimlund linkage rule (described on pp. 208) can result in a sample size as large as the entire population. It is evident, then, that this linkage rule should only be used when there is an immaterial error and, therefore, the maximum sample size should

TABLE 44

STRATEGY 7: TABLE 38 SAMPLE SIZE STATISTICS

	Environments		
	E1	E2	E3
Distribution of NSMPL			
Mean	82.7933	196.4917	205.3500
Standard deviation	25.6627	36.9147	34.0311
Skewness	.83	-.14	-.27
Kurtosis	.50	-1.12	-1.66
Maximum	183.0000	237.0000	237.0000
Minimum	60.0000	103.0000	103.0000

TABLE 45

STRATEGY 8: TABLE 39 SAMPLE SIZE STATISTICS

	Environments		
	E1	E2	E3
Distribution of NSMPL			
Mean	61.8733	209.9217	221.2167
Standard deviation	17.7965	61.0441	48.1013
Skewness	9.64	-1.90	-2.84
Kurtosis	92.09	1.80	6.37
Maximum	237.0000	237.0000	237.0000
Minimum	60.0000	60.0000	60.0000

be established independently.

On the other hand, the minimum sample size requires use of the Bayesian framework of the Felix-Grimlund model because other prior information (i.e., internal control information) is incorporated in the analysis. The use of a Bayesian framework means that the strength of the prior information determines how much substantive testing is carried out. Hence, it is possible for the substantive test sample size to even be as low as zero. Although in actual practice such extreme reliance is disallowed for, it was felt that in the interests of statistical performance of the model, those sample sizes predicted by the Felix-Grimlund linkage rule should be used in the simulation. This does however, introduce some noncomparability between the purely Bayesian Felix-Grimlund use of internal control information and the non-Bayesian (or quasi-Bayesian) methods. Nevertheless, establishing a non-zero minimum substantive test sample size appears to bring even more problems and so it appears best to just let the mathematics of the model decide on the optimal sample size. This is what was done in the simulation and, thus, the range of sample sizes for the strategies using the Felix-Grimlund model is from 0 to 120. Table 46 presents the performance of the Felix-Grimlund model at the maximum sample size. It is followed by table 13 to facilitate comparison of performance with the conventional DUS for the same sample size. Two major aspects of the Felix-Grimlund model stand out in making the above comparison. First, the Felix-Grimlund model is not conservative in the sense that it overstates the probability of

TABLE 46

PERFORMANCE OF STRATEGY 15: E - C

(Informationless, no linkage with
Felix-Grimlund substantive test)

Environment	Planned α	Actual α	Actual β	Planned β	Actual Confidence Level	Average Sample Size
E1	None directly	.000	.000	.05	No basis for Computing	120
E2	None directly	.156667	.000	.05	"	120
E3	None directly	.939999	.000	.05	"	120
E4	None directly	.000	.073333	.05	"	120
E5	None directly	.000	.000	.05	"	120
Realistic Average		.187999	.001833			120
Conservative Average		.219333	.014666			120

TABLE 13

PERFORMANCE OF STRATEGY 14: E - B

(Informationless, no linkage with DUS)

Environment	Planned α	Actual α	Actual β	Planned β	1-Actual Confidence Level	Average Sample Size
E1	None directly	.000	.000	.05	.000	120
E2	None directly	.400	.000	.05	.000	120
E3	None directly	.989999	.000	.05	.000	120
E4	None directly	.000	.013333	.05	.000	120
E5	None directly	.000	.000	.05	.000	120
Realistic Average		.2685	.000333			120
Conservative Average		.278	.002666			120

accepting an account with a material error thus minimizing the risk of Type II error. This is evident from looking at the actual β risk in E4 of table 46, which although it is not considered significant by the rule given on p.237, it is higher than for the other two substantive test methods.

To further analyze the actual β risk level in E4, several assessments were made of this risk each the result of 600 different samples and statistical tests:

TABLE 47

FURTHER ASSESSMENT OF FELIX-GRIMLUND β RISK
(Planned β = .05)

Assessment Number	Actual β Risk for E4
1	.071667
2	.060
3	.063333
4	.056667
5	.045000
6	.063333
7	.056667
8	.060
9	.065
10	.048333

Thus it is evident that Felix-Grimlund does result in a somewhat higher than nominal β risk. This is either the result of the fact that the normal distribution assumption of error sizes per dollar is not sufficiently robust as Felix and Grimlund have argued; or that the extended beta approximation to the beta-normal distirbution is insufficiently accurate.[15] Either way, the model in its present form comes nowhere near to containing the β risk as well as conventional DUS.

On the other hand, and this brings up the second major feature of the Felix-Grimlund model, the model proves to be very effective in keeping the α risk at a minimum. This is particularly noteworthy of E2 performance where the α risk is less than half of that for either DUS or STMPU. Thus there appears to be considerable potential for this model if only the β risk could be reduced for the exactly material amount of error. Perhaps one approach would be to slightly reduce the acceptable probability of a material error from .05 to, say, .045. This is consistent with the kinds of adjustments advocated by Don Roberts and is also the method used to make the Mann C value more suitable in linkage II(B). To the researcher this appears to be a reasonable way of dealing with statistical procedures which, except for a predictable bias which can be adjusted for, perform satisfactorily for audit purposes.

Another possible way of dealing with the Felix-Grimlund model's failure to keep actual β risk within the planned level is to combine it with conventional DUS (since both samples are taken the same way) and set up special rules to deal with those situations where the two

---

[15] See Felix and Grimlund, pp. 33-34.

results disagree. It is certainly worthwhile to explore various ways of improving the Felix-Grimlund substantive test method. Considering the enormous reduction in α risk the model promises for various degrees of immaterial errors (which according to the Clarkeson, Gordon & Co. occur on 95% of all audits) and only a slight amount of excess risk at the exactly material amount of error; the model has the potential of becoming the best substantive testing method in auditing.

With the impact of internal control information, the model becomes a Bayesian one in the formal way it integrates the information. That is, a non-diffuse prior distribution is constructed before any substantive testing takes place. In the omniscient case this prior is given a very high weighting of $n^* = 750$, and then the procedure given in pp207-8 and appendix V is used to compute a substantive test sample size. This size substantive test is then taken and a statistical decision is reached on accepting or rejecting the book value total. The performance of this omniscient strategy is given in table 48.

Thus the internal control hypothesis is being upheld by the Felix-Grimlund model also. Note that the realistic average sample size works out to be the smallest of any strategy. Of course, this is due to the fact that qualitative and institutional factors have not been considered in the Bayesian analysis and, hence, the sample sizes are not strictly comparable to the non-Bayesian ones.

One might wonder why the sample size for E3 proves to be the maximum allowable instead of something smaller considering that it still

TABLE 48

PERFORMANCE OF STRATEGY 6: ADC

(Omniscience, Felix-Grimlund Linkage & Substantive Test)

Environment	Planned α	Actual α	Actual β	Planned Combined Risk	Average Sample Size
E1	None directly	.000	.000	.05	0
E2	None directly	.000	.000	.05	0
E3	None directly	.939999	.000	.05	120
E4	None directly	.000	.733333	.05	120
E5	None directly	.000	.000	.05	120
Realistic Average		.02350	.001833		24
Conservative Average		.187999	.014666		72

is, essentially, immaterial. The reason for this is that the mathematics of the sample size computation is such that it cannot distinguish between E3 and anything over a material error for a prior weight of $n^* = 750$. In fact it cannot distinguish E3 from E4 even when the weight is as high as $n^* = 2000$. Turning the problem around, it was found that the Bayesian auditor put complete reliance (i.e., planned substantive sample size of zero) at about $R = .915$ but no reliance for any reliability less than that. Thus in certain respects the Felix-Grimlund model behaves very much like the II(B) linkage using the C value computed by the Mann method. That is, both methods result in extremes of reliance for very tiny changes of the estimated $R = 1 - \rho$ value, and they both tend to have excessive actual combined risks due to unwarranted reliance effects. It appears that this is the result of computing and basing decisions on the probability of a certain value occurring as opposed to constructing a decision rule based directly on measures of the amount of error.

Turning to the objective case of internal control information where the strategies are driven by actual compliance test results, the performance as indicated in table 49 results.

As expected, due to unwarranted reliance the average β and combined risk climbs to .08 which is significantly (i.e., greater than .025) beyond the nominal level of .05. However, the increase beyond the informationless case is only .006667 = .080 - .073333. Thus most of the excess β risk is due to the Felix-Grimlund substantive test procedure and not to the linkage rule or to the sampling error

TABLE 49

PERFORMANCE OF STRATEGY 12: CDC

(Objective, Felix-Grimlund
Linkage & Substantive Test)

Environment	Planned α	Actual α	Actual β	Planned Combined Risk	Average Sample Size
E1	None directly	.000	.000	.05	96.63
E2	None directly	.000	.000	.05	.000
E3	None directly	.910667	.000	.05	114.68
E4	None directly	.000	.080	.05	115.88
E5	None directly	.000	.000	.05	120.00
Realistic Average		.136600	.002		71.41
Conservative Average		.182133	.016		89.44

associated with the internal control information.

The relatively large average sample size for E1 arises because these sample sizes are required to fulfill the mathematical needs of the Felix-Grimlund model as explained on pp.207-8. These sample sizes appear anomalous compared to the zero average for E2 because the error rate in E1 is not sufficiently large for the prior sample size weight of n* = 250 to meet the mathematical needs of the model. Grimlund has rationalized this weakness by arguing that it is unlikely a formal modeling would be necessary for an E1 type situation (see footnote 70). But certainly this is a deficiency of the model and highlights the problems associated with insufficient error observations and data availability issues discussed in chapter two. Thus, for the objective case the model ends up with average sample sizes greater than the non-Bayesian ones.

Note, however, that the α risk for E1 is zero and β risk for E5 is zero thus indicating that away from the materiality threshold the model performs fairly well. Therefore, the strategy does show promise for improving on non-Bayesian performance if some valid rules could be devised for obtaining minimum sample sizes and dealing with the slightly excessive β risk in E4. Tables 46 and 48 certainly show there exists potential for improving on present auditing methods. However, it is not a goal of the dissertation to develop such an improved model.

Finally, to facilitate comparisons of objective case strategy performance, table 50 reports the statistics for the substantive test sample sizes when they varied as a result of the varying compliance

TABLE 50

STRATEGY 12: TABLE 49 SAMPLE SIZE STATISTICS

	Environments		
	E1	E3	E4
Distribution of NSMPL			
Mean	96.6333	114.6833	115.8833
Standard deviation	34.0234	22.4739	19.0129
Skewness	-1.11	-4.54	-5.08
Kurtosis	-.36	19.64	25.75
Maximum	120.0000	120.0000	120.0000
Minimum	20.0000	.0000	.0000

test results; and table 51 provides statistics on the four moments of the estimated beta-normal distribution.

This completes the report on the performance of the audit sampling strategies.

5.5 Summary

The preceding sections report on the performances of various audit strategies which have been proposed or are actually used in practice to reach a conclusion about the accuracy of the book values in various accounting environments. Based on the available evidence, these environments appear to be representative of what auditors actually encounter in the field. Thus it appears that the strategy performance should be representative of what would take place in practice. Bearing in mind the cautions expressed in evaluating a simulation study as discussed in chapter one and recognizing the need for future research to corroborate these findings; it, nevertheless, appears reasonable on the basis of present evidence to reach the following tentative conclusions.

Major findings (related to the
major goals of the dissertation)

I. The statistical validity of the internal control hypothesis of auditing has been confirmed for both the "weak" and the "strong" forms of the hypothesis. In general the hypothesis holds for all linkage rules and substantive testing methods considered. However, the more general internal control hypothesis of auditing does appear to be conditional on the amount and quality of the internal control information available. For example, in the extreme case of subjective

TABLE 51

STATISTICS FOR THE FELIX-GRIMLUND MODEL
AT SAMPLE SIZE 120

Results of 600 Simulations:	Environments E1	E2	E3	E4	E5
Distribution of Estimates of Mean of Beta-Normal					
Mean	.34372+005*	.32350+006	.67861+006	.69335+006	.10280+007
Standard deviation	.46035+005*	.91698+005	.19548+006	.19565+006	.28074+006
Skewness	1.2	.7	.6	.4	.8
Kurtosis	.5	.3	-.3	-.5	-.3
Maximum	.19015+006	.64498+006	.12793+007	.12104+007	.18019+007
Minimum	.00000*	.13205+006	.33307+006	.31303+006	.55240+006
Distribution of Estimates of Variance of Beta-Normal					
Mean	.12062+010	.16645+011	.33934+011	.34892+001	.49714+011
Standard deviation	.29019+010	.60920+010	.11798+011	.12016+011	.17183+011
Skewness	2.2	.7	.5	.3	.7
Kurtosis	3.3	.2	-.4	-.6	-.4
Maximum	.11654+011	.38481+011	.67630+011	.64364+011	.97017+011
Minimum	.00000	.39863+010	.10000+011	.10000+011	.10000+011
Distribution of Estimates of Third Central Moment of Beta-Normal					
Mean	.14506+015	.16381+016	.29850+016	.30864+016	.39141+016
Standard deviation	.36349+015	.77253+015	.11625+016	.12049+016	.15150+016
Skewness	2.4	.8	.4	.2	.3
Kurtosis	4.2	.7	-.5	-.6	-.8
Maximum	.13804+016	.47294+016	.66184+016	.66184+016	.74954+016
Minimum	.00000	.10000+011	.10000+011	.10000+011	.10000+011

TABLE 51--Continued

Results of 600 Simulations:	Environments E1	E2	E3	E4	E5
Distribution of Estimates of Fourth Central Moment of Beta-Normal					
Mean	.55524+020	.11719+022	.42029+022	.44285+022	.86282+022
Standard deviation	.14546+021	.81839+021	.27814+022	.28111+022	.58624+022
Skewness	2.6	1.4	1.1	.9	1.1
Kurtosis	5.4	2.6	.7	.3	.5
Maximum	.64408+021	.52349+022	.14206+023	.13100+023	.28640+023
Minimum	.00000	.10000+011	.10000+011	.10000+011	.10000+011

*The mean, standard deviation, maximum, and minimum values are written in logorithmic notation with the integer after the (+) sign representing the power of 10 that the decimal to the left of the + sign is multiplied by (e.g., .34372+005 = .34372 x 10^5 = 34,372).

information through strategy 16 of table 34, it is evident the hypothesis does not hold for the most efficient linkage rule and substantive testing method. (Compare table 34 to its objective case counterpart, table 27.) Even for lesser amounts of uncertainty (e.g., based on sampling risks only), there may be significant increases of β risks beyond the "informationless" level (but not significant as used here beyond the nominal level); for example, compare table 36 to table 38 or table 37 to table 39. On the other hand, some linkage rules and substantive test combinations do not result in significant risk increases with imperfect information, for example, compare table 13 with table 27 or table 28. Thus the validity of the general internal control hypothesis is dependent on the quality of internal control information, the type of linkage used, and the substantive testing method; with the conservative linkage expected to perform better as the quality of internal control information declines. (Quality, in turn, is determined by the internal control model, the amount of compliance testing, and the expertise of the participating auditor.)

II. Generally, it is statistically valid to allow the nominal β risk to get as high as .5 and still control audit risks close to or within the nominal level. (See comparisons in finding I.) This means that with good internal control information it is statistically valid to reduce the smaple size for substantive tests as much as 75% when using STMPU and DUS, and perhaps more in a purely Bayesian framework. These upper limits represent the upper bound on the value of internal control information in terms of the savings in substantive tests when there is an immaterial amount of error.

The fact that the sample sizes with complete reliance proved to be the smallest they could possibly get in a realistic situation and still provide statistically valid estimates, indicates that these results should continue to hold when the sample sizes are larger.[16]

III. A comparison of the linkage rules indicates that the Elliott and Rogers linkage performs about as well as the SAS No. 1 Sec. 320 linkage for both DUS and STMPU when the internal control information is "objective" (compare table 27 to 28, and table 38 to 39). However, SAS No. 1 is distinctly superior in the limiting omniscient case or when extremely good information about internal controls is available (compare table 36 to 37 or table 24 to 26). This is due to the conservative tendency of the gradualism expressed by the Elliott and Rogers linkage. However, such superiority disappears and is even reversed under less than perfect internal control information and use of STMPU.

One possible advantage of the SAS No. 1 linkage is that the auditor is not required to identify rather arbitrary different degrees of reliance and relate them to grades of internal control; instead, using the SAS No. 1 linkage he just needs to identify the exactly

---

[16]The substantive test sample sizes used here are smaller than that of earlier studies. It is unlikely STMPU could have been made much smaller considering that with a sample size of 60, there are only three observations in each stratum. Similarly, the smallest DUS sample sizes are at the low end of the range reported in examples.

Since the estimates generally perform in conformity with sampling theory, larger smaple sizes will tend to be even more in conformity. Larger sample sizes can result by using a higher confidence material error rate and make a conservative downward adjustment in computing the C value using the Mann method.[17]

The Clarkeson & Gordon linkage is, as expected, distinctly sub-optimal in utilizing internal control information for the objective and omniscient cases (compare table 27 to 26 or 24, and tables 29 to 28 or 27), but this conservatism can prove to be optimal under sufficient conditions of uncertainty about internal controls. (Note that with the introduction of compliance test sampling error in strategy 11, the combined risk is still the same as in the omniscience case--compare table 21 with table 18.)

The Felix-Grimlund linkage has a potential for being superior to all others (compare table 48 to 46 and average sample sizes of table 48 to those of table 26), but the Felix-Grimlund substantive test method has certain technical problems (note the high β risk) in its presently developed form which limits the usefulness of the associated linkage rule.

IV. When too many judgmental errors enter the process of determining the extent of substantive testing, strategy 16 arises (table 34) comparing its performance with its objective case counterpart in table 27 shows that the main effect of judgmental errors appears to be an increase in the actual combined risk to twice the nominal level and

---

level and/or a greater precision.

[17]This procedure is explained on p. 258-260.

three times the objective level. Certainly this risk could have been reduced by increasing the average substantive test sample size. Hence, it is not difficult to understand why most linkage rules used in practice tend to be conservative (inhibit reliance on internal controls) and perhaps why many auditors do not appear to bother relying on internal controls (as indicated by behavioral studies).

Considering the large savings available from internal control information (See conclusions I and II), it appears that auditors should be encouraged to make use of it and trained to do so. The benefits can include not only a significant reduction in substantive test sample size, but, in the case of STMPU, a significant reduction in actual α risk as well. (Compare table 37 to table 15.) In fact this means internal control information can be more valuable when STMPU is used.[18]

V. The relative performance of DUS and STMPU are perhaps contrasted most sharply by comparing table 9 with table 15 and noting that in table 11 DUS has a sample size almost half that of STMPU in table 13. Both strategies have sample sizes computed as they would be in practice given the goals of the simulated auditor. These results essentially corroborate what the DUS advocates have argued for years--that DUS is generally superior to the classical sampling approaches for use in auditing.[19] In particular the false alarm (α) risk is less upto at least 50% materiality. Thus the supposed excess false alarm

[18] This is also true of the associated linkage rule.

[19] The researcher also confirms that DUS is a much easier procedure to program and apply in practice.

risk of DUS for few errors does not exist when the TACS evaluation method of Teitlebaum is used.

The Felix-Grimlund model almost outperforms DUS except for the somewhat excessive β risk associated with the method. (Compare table 46 to table 13.) However, as indicated on p.258-260 there exist ways of reducing this β (with an undetermined increase in α risk) and a more thorough theoretical analysis could probably devise a more attractive solution to the problem; so it appears that this model has a very high potential for becoming the best single substantive test method for use in auditing.

Lesser findings

VI. The crude method for computing lower bounds on system reliability appears to be too conservative for use in auditing (at least when five or more internal control components are involved), and so the Mann method was used in the simulation. The Mann method, although somewhat optimistic (and thus resulting in an increase in unwarranted reliance, and, ultimately, in Type II errors as reflected in the "objective" strategies), proved to work fairly well for audit purposes. For example, compare table 36 to table 38, table 24 to table 27, and table 27 to table 34. Thus the Mann method appears to have good potential for use in auditing for evaluating internal controls by integrating the results of several compliance tests (particularly tests of key internal control procedures which naturally result in a series system structure as discussed in chapter two and three). Considering the results of prior research with the Mann method, perhaps much of the optimism of the method would be reduced when there are fewer components (key controls) in the system.

VII. The validity of using the confidence level associated with the compliance tests as the C value in the SAS No. 1 Sec 320B formula (as argued on pp.193-4) appears to be confirmed by the simulation (e.g., compare table 37 to table 39, or table 26 to table 28). This is the first known success in using the SAS No. 1 formula directly, and thus the significance level appears to be a good objective surrogate for the probability of the existence of a material error. Of course, this depends on how accurately the auditor can establish the relationship between compliance errors and total dollar error impact on the system.

Actaully, the significance level associated with the exactly material monetary error rate (1-R) = .9 was not used in the simulation due to the bias of the model used--the nominal confidence overstates actual confidence as discussed on pp. 262 . However, this deficiency is only a problem for the II(B) linkage and adjustments were made accordingly. The adjustment does not imply that using the significance level for C is inappropriate, rather it counteracts the bias associated with the approximation provided by the Mann method for bounds on system reliability.

VIII. It was found that the procedure of increasing basic precision for low error rates when planning for sample sizes using DUS is inappropriate because this practice tends to increase the α risk associated with the substantive test (See table 35). In fact it appears that an optimal precision (relative to STMPU performance) is to use 1/2 the materiality level. This implies that for any confidence level the planned smaple size should be twice the discovery sample.

The general rule that appears to underlie all the sampling strategies is this: Do not reduce sample size unless the associated confidence level is also reduced (according to the sample size formulas) or an equivalent adjustment is made (e.g., under the positive approach for STMPU, the adjusted precision rather than the achieved precision should be used in constructing the decision interval).

IX. There is no valid positive approach for STMPU in auditing (in the sense nominal risks $\geq$ real risks) except in the case where it works out to be the equivalent of the negative approach (i.e., use adjusted instead of achieved precision).

X. The complement of the actual confidence level is a poor indicator of actual α and β risks of a statistical estimator for various degrees of error. Hence it is not surprising that earlier studies were were not able to reach a conclusion about the relative performance of estimators having reliable nominal confidence levels. See tables 14, 34, 40, 41, 42, and 43 for good examples.

This then concludes the summary of strategy performance. The next next chapter discusses the implications of these findings for prior and future research, and for present practice.

CHAPTER SIX

Summary and Conclusions

6.1 Summary of the implications of the findings for audit practice

The general conclusion that can be reached from the study described here is that the statistical validity of the internal control hypothesis has been confirmed even when internal control reliance results in the smaller substantive test sample sizes that are likely to be used in an actual audit. However, statistical validity does not mean that the hypothesis holds when nonsampling errors are introduced to the audit process. Thus, in practice, if there is too much error in auditor judgment the hypothesis may still not be valid. Nevertheless, empirical evidence that in a purely statistical sense audit process reliability can remain constant or even be improved in actual auditing environments while reducing the extent of substantive testing is important for developing a scientific basis for auditing theory. At the very least, then, this study has begun establishing the conditions under which the hypothesis is valid.

Although prior statistical work in this area was certainly appropriate in a statistical sense, it did not provide the relevant measures of audit strategy performance that are necessary for auditors to make decisions about these strategies. In particular the measures were not discriminating enough to choose between the two major substantive test methods, DUS and STMPU. Of course, much of this inability to discriminate may have been due to the particular environmental circumstances chosen but the results of this study indicate it is much more likely that the measures used in prior research are the main reason these relative performances issues had not been resolved.

The present available evidence on audit environments indicates that the environments used in the simulation contain characteristics which are commonly found on actual audits. Of course, this may still be a debatable point for some until more evidence is brought forth on auditing environments. On the other hand, there is little room for argument that the auditing environments considered here do represent a large and important class of actual accounting situations which are important in their own right for assessing the performance of audit sampling strategies. At the very least then it is possible to reach some practical conclusions concerning audit strategy performance for this class of audit environments.

For the environments represented here, which may or may not represent a majority of populations for which substantive testing is done, but which do represent a significant portion of such populations; it is evident that DUS is a clearly superior substantive test procedure to STMPU. This is not only true in terms of actual α and β risks but also in terms of the validity of the internal control hypothesis of auditing. It is "safer" to use internal control information with DUS due to its inherent conservatism in controlling the combined and β risk. That is, it is less likely with DUS that the use of a particular level of internal control information and linkage rule will result in actual combined risks greater than nominal as a result of internal control reliance. Since DUS conservatism holds for all possible environments and also has equivalent α risks for a large class if not a majority of environments, there appears to be little room for argument that it is superior to STMPU for audit use. One possible area of future audit research then is to identify the conditions under which STMPU has lower α risks (for the same sample size) than DUS and to establish how frequently these conditions occur in audit practice.

The practical implications of the dissertation for linkage rules and value of internal control information are less clear cut. As indicated in the conclusion of chapter five, the validity of the internal control hypothesis, and hence the validity of the linkage rules and value of internal control information, is dependent upon the quality of the internal control information. If there are relatively few non-sampling errors (how much is dependent on the results of future research) and the sampling errors of compliance tests are relatively low (95% confidence appears to be sufficient), then there is a great deal of value (in theory almost a 75% reduction in total substantive testing over all audits using a "realistic" averaging) to internal control information.

It is premature to make general statements about the performance of linkage rules other than the conditional ones already made at the end of chapter five. However, some general guides which may be of value to practitioners can be stated. First, a substantive sample size reduction as a result of reliance on internal controls should be accompanied by a confidence level reduction as determined by the basic sample size formulas used for planning purposes. Since most linkage rules work by first computing the lowered confidence level and then the accompanying sample size, this first rule is not violated in such situations. However, any other effective reduction in sample size, such as by increasing planned precision (either in DUS or STMPU), can lead to increased α risks.

Second, conservative linkage rules, which reduce the frequency with which internal controls are relied upon or reduce the sample size reduction of the substantive test, generally reduce the value of the internal control information. Thus although the conservatism of the linkage rules reduces the actual combined risk of an audit strategy, this may be done at the cost of making the internal control information less cost-benefit justified. On the other hand, conservative linkage rules are more likely to make the internal control hypothesis valid in practice by maintaining the audit risk levels when there is reliance on internal controls.

Third, use of the statistical significance level appears to be a valid way of obtaining a more objective measure of "reliance" on internal controls. This is so because it requires the auditor to only make an assessment of the exactly material compliance or monetary error rate for the system and avoids the problem of evaluating grades of internal control. In particular the significance level associated with a set of compliance tests for an internal control system can provide a valid and objective assessment of the "C" value in the basic SAS No. 1 Sec. 320 B linkage formula.

Fourth, in terms of maximizing the value of internal control information the following ordering of linkage rules holds as the quality of internal control information declines: (1) SAS No. 1 linkage, (2) Elliott and Rogers linkage, and (3) Clarkeson, Gordon, & Co. linkage (or equivalent for STMPU). This ordering indicates which of the three non-Bayesian linkage rules performs best as the quality of internal control information declines. That is, SAS No. 1 performs best when the quality is very good and the auditor feels he is knowledgeable about the internal control system; Elliott and Rogers linkage can perform better in maintaining risks when there is more sampling error concerning system reliability and the auditor feels there may be some error in his interpretation of system reliability which can be compensated by the conservatism of the rule; and the Clarkeson, Gordon, & Co. linkage can be optimal when there is significantly more sampling and nonsampling error in evaluating system reliability than under the conditions of the first two linkages. All linkages are safest (in minimizing combined risk) when used with DUS.

These are obviously very nebulous guides but, on the other hand, so is a measure of the auditor's knowledgeability of the internal control system. Until more is known about auditor judgment processes and their correlation to real world relationships, more precise guidance would be premature.

Fifth, auditors ought to be aware that conservatism of the linkage rule is not the only way of controlling for combined risk. Other ways of controlling this risk is to more conservatively model the internal controls and/or evaluate the compliance tests conservatively. An example of a conservative evaluation of compliance test results is to use the crude method for obtaining a lower bound on system reliability. These procedures can be used to introduce conservatism rather than through the linkage rule itself. Thus it may be that the SAS No. 1 linkage performs well even with significant judgment errors when conservatism is introduced thru the modeling of the internal controls--but this is just speculation at present and requires further research before more precise guidance can be provided.

Sixth, auditors should not use the pure positive approach with STMPU. The achieved precision should always be adjusted as indicated in chapter four in order to control the sampling risks at the planned level.

Finally, the eventual issue of whether internal control evaluation is justified is dependent on the relative costs of obtaining this information or extending the substantive tests. As indicated in chapter one, this depends on the cost structure applicable to the particular firm. No attempt is made here to justify any particular cost structure. Instead the "conservative" and "realistic" averages of the key aspects of strategy performance, α risks, β risks, and substantive test sample size have been provided to assist in making this evaluation.

Besides the empirical evidence, the practitioner should also find useful the theoretical arguments presented for assuming a series structure for most internal control systems and the methods for obtaining conservative evaluations of system reliability. These have been presented in chapters three and four and appendixes I, II, and IV.

6.2 Relationship of the study to prior research

In relation to prior auditing research this study has (1) demonstrated the value of more accurately modeling internal controls thus justifying prior and future research attempts in this area; (2) shown that direct measures of actual α and β risks provide important additional information for discriminating between strategy performance which was not possible on the basis of previous studies; and (3) shown that the validity of the internal control hypothesis in practice is limited by the extent of judgmental errors in practice and not the validity of the statistical test under audit conditions.

6.3 Implications for future research

What then are the implications of the dissertation for future research?

It appears that the research effort can be extended usefully in two major directions. First, more statistical research is needed: there is a need to assess audit strategy performance in different accounting environments particularly extreme differences from the ones considered here. For example, use of less skewed book value distributions and/or dollar error sizes not proportional to book value can be used to determine whether and under what conditions STMPU performance improves relative to DUS. Another variation of this direction of research is to simulate the sensitivity of audit strategy performance to particular kinds of judgmental errors in the audit process. For example, linkage rules or auditing environments can be changed to see what effects such things as failure to identify all sources of error in an internal control system or failure to correctly establish the relationship of compliance to monetary errors has on audit strategy performance. This type of simulation research can be particularly useful when combined with behavioral research results on these topics. And, of course, there is the need to further empirically and theoretically analyze the new methods proposed here, the Mann method and Felix-Grimlund model before they can be made fully operational for use in practice. The research results here indicate both methods have a very high potential for use in auditing and, therefore, they deserve more attention in future research.

The second major direction of research indicated by this study is behavioral. The second auditing standard of fieldwork assumes that an auditor has sufficient expertise to make use of internal control information on an audit. This does not mean that judgmental errors do not arise in evaluating internal controls or determining the extent of substantive tests. Rather, the implied assumption is that such errors are not great enough to invalidate the internal control hypothesis. This is an empirical question for which the present evidence is ambiguous considering the behavioral studies reported in chapter two. It is thus important that behavioral research be directed to isolating the judgments in the basic audit process for which there exist significant error potential, and estimating the distribution of such errors. The kinds of errors which arise have an important bearing on the practical validity of the internal control hypothesis and in affecting auditor training programs. For example, if it is found that large judgmental errors occur at the linkage stage, then the results of this study indicate that training in the application of the linkages used here should significantly improve auditor performance in terms of efficiency and sampling risks. Similarly, if it is found that large judgmental errors occur in the internal control evaluation stage, then use of more formal methods such as the reliability model and evaluation techniques proposed here may considerably aid auditor performance. At the very least the behavioral research on auditor judgment processes will help clarify where practice now stands so that theoretical models can be based on more accurate assumptions.

Thus it is evident from this discussion that there remains considerable need for further research before many of the remaining practical issues concerning choice of audit strategy can be resolved. Nevertheless, to this researcher it appears that reasonable evidence on audit strategy performance can soon be provided for basing such decisions. Much depends on progress in the research extensions just discussed. This research is important for putting auditing theory and practice on a firmer scientific basis. The researcher has already begun extending empirical simulation work in the statistical research direction and in the future he hopes to work with others in the behavioral research direction.

APPENDIX I

Additional Modeling Issues

The purpose of this appendix is to attempt to clarify some of the abstraction issues raised in chapter three.

The basic goal is to explore and compare the implications of the system or environment as described in chapter three (henceforth to be called the simple system) to the implications associated with attempting to simulate a more complex (more "realistic") system. It is hoped that by this means it is possible to show that under certain conditions the application of audit strategies to the two systems is equivalent and that therefore the results of the simulation for the simple system would be at least as generalizable as it would be for the complex system. In addition, by going through this exercise the researcher hopes to make clear that the complex system has to make more rather arbitrary assumptions (arbitrary because there does not appear to be any empirical evidence available on these matters) about the accounting environment and this may ultimately reduce rather than increase the generality of the results that are obtained.

The approach of the rest of this appendix is to first describe the more complex system in detail and consider the new problems introduced. The implications of the complex system for the simulated audit strategies are then considered and compared to the implications of the simple system. The appendix concludes with a summary of the results of the analysis and consideration of some related issues involving the goals of the research project. The reliability approach, for reasons discussed in chapter two, is assumed in modeling using both the simple and more complex system representation.

At the outset let it be recognized that this is not to be a simulation of an entire audit, but rather of only one item in the financial statements. More specifically, it appears that the controversy centers around using the simple system to represent an account balance as opposed to a line item (it is fairly evident that the simple system could reasonably represent an internal control system for line items or transactions). Therefore, to be specific, consider what appears to be a more realistic representation of a trade accounts receivable file.

It is possible to start by considering what would perhaps be an ideal situation from the auditor's standpoint. Figure 10 illustrates an accounts receivable file arrangement whereby all the transactions pertaining to a particular account in the file and all the pertinent attributes relating to the transactions and the account are provided to support the validity of the aggregated total of the account balance outstanding at any particular point in time. Each account (designated by a unique account number), therefore, is supported by a complete file of all transactions related to it. The book values and audit values are as defined in chapter two. The transactions are chronologically ordered from the most recent (TR#1) to the beginning of the period (or earlier?). The entry designation indicates whether the transaction resulted in a debit or a credit to the account balance.

FIG. 10

COMPLEX SYSTEM REPRESENTATION

ACCOUNT NO.		BOOK VALUE	AUDIT VALUE

One trade account receivable with an outstanding balance at year end supported by a file of transactions.

File of transactions supporting the book and audit value at year end for the account receivable balance.

type	transaction #	entry	attributes									
	TR #1	Dr	X	X	X	X	X	0	0	0	BOOK VALUE	AUDIT VALUE
	TR #2	Dr	X	X	X	X	0	X	X	X	BV	AV
	TR #3	Cr	0	0	0	X	X	X	X	X	BV	AV
	TR #4	Dr	X	X	X	X	X	0	1	0		

The attributes section theoretically includes all attributes relevant to the internal controls associated with the account's balance. (Note, however, that the attribute of correct summarization of all transactions is really an attribute associated with the aggregate account balance only and cannot be associated with any individual transaction--by the way this is considered a separate audit objective by Arens and Loebbecke, p. 202, and as a consequence is an important attribute.) Not all attributes apply to all transactions--the "X" in the attributes column indicates that there is no such attribute for that type of transaction. The "type" column for transactions more clearly defines the kind of transaction processed (e.g., recognition of sale on credit, receipt of cash payment, and recording of sales returns and allowances).

The reason for generating this transaction file is so that the sum of all transaction book values and audit values will equal the ending balance book and audit values, respectively, (assuming no errors in addition) of the trade account receivable. That is, much more information about the internal controls applying to the transaction processing stream during the period is made available. Unfortunately, there are major problems associated with generating such a transaction stream and these are discussed further in analyzing the system and what information is available to auditors.

The first question which might arise with simulating the above system is: How far back in time is it necessary to construct a transaction file for the account? Theoretically, one could conceivably go back to the inception of the firm or at least to the opening of the account--in fact this appears to be the only approach if one wanted to work only with transactions. Call this possibility case 1.

A more plausible approach would be to consider going back only as far as the previous audit and use the beginning balance of the period as the starting value, with all current period transactions (debits cancel credits) added on to get the ending balance. Call this case 2.

Yet another possibility is to recognize only the transactions outstanding at the end of the period and that have not yet been closed with the customer (e.g., all unpaid sales invoices or invoices still in dispute). In fact this approach appears the most logical since it restricts attention only to those transactions (and related attributes) relevant to the outstanding balance at the end of the year. Call this case 3.

A problem with case 3 is that it would not result in an accurate depiction of compliance testing as appears to be typically done by the auditor. That is, auditors attempt to test transactions throughout the year instead of concentrating on particular subperiods, such as year end, which would be implied by a case 3 representation. (See pp. 202-203 in Arens and Loebbecke for more details.)

So now the further specification of the complex system is continued by assuming case 2 holds. (It is assumed case 1 is impracticable because it essentially requires knowledge of the entire history of internal control operations although this may be the only way to adequately account for "residuals" or prior period errors affecting this period's book values. Case 3 is considered in more detail later.) This will require additional discussion before getting to the assumptions underlying construction of such a system.

First it must be recognized that when auditors test for compliance, they do not do it from one master file of transactions all grouped by customer. Instead, there are typically several files of transactions (e.g., sales invoices, remittance advices, and credit memo's) which are usually in chronological order (based on a prenumbered sequence) and are therefore mixed with similar transactions of other customer accounts. These files can be classified into two categories, transactions which result in a debit to the accounts receivable and transactions which result in a credit entry to accounts receivable. However, the significance of this distinction appears negligible because errors in both directions can occur in both categories and hence either category can produce an over-or under-statement in the accounts receivable balance. It appears, therefore, that the distinguishing feature of each transaction file is not so much which category it belongs to, but rather the relevant attributes which can indicate sources of dollar errors within each transaction file.

The more "realistic" complex system should therefore be represented as in figure 11 which follows.

Note this system consists of two general components designated A and B in figure 11. "A" consists of the file of records representing the accounts receivable outstanding at year end while "B" consists of

FIGURE 11

MORE VALID COMPLEX SYSTEM REPRESENTATION

A. File of Trade Accounts Receivable as of Year End

ACCOUNT NO.	BOOK VALUE	AUDIT VALUE	

B. Transaction Files for a Year

File of Sales Invoices | File of Remittance Advices | Other Transaction (Line Item) Files

several transaction files supporting the values contained in the "A" file. Ideally, the auditor would like to test the transaction file after year end (see, for example, p. 202 of Arens and Loebbecke). However, the auditor frequently resorts to interim date testing only and then tries to obtain assurance that the system did not change radically afterwards.

It is now possible to consider the additional assumptions (that is, in addition to those already made for the simple system) that need to be made in simulating this complex system.

1. One must decide how large each transaction file should be.

2. One must decide how many transactions in each file apply to each account balance (and to account balances which are no longer outstanding).

3. One must decide how the book and audit values are to be distributed for each account in each file. Note this would be a very difficult operation because after these values are randomly generated for the entire year (and appropriately distributed chronologically), one must assure that the algebraic sum totals to the ending balance for each particular account (assuming there are no summarization errors errors). This would entail considerable computational effort because there are over 7,000 individual accounts for which this would need to be done.

4. One must decide on a beginning of the year book and audit value for each account. (These are necessary so that the sum of the transactions and the beginning values equal the ending values [assuming no summarization errors]).

5. Since the transaction files attributes indicate the performance of internal controls over time, one must decide how the compliance error rates vary over time. This assumption would probable prove to be the least troublesome to justify since a good internal control subsystem would not be expected to fluctuate wildly during a period. If it did the auditor ould feel so unsure he probably would not want to rely on the system. Thus error rates would not only have to be sufficiently low but sufficiently stable as well. The auditor would be particularly concerned with a situation where the compliance error rates are low in the period preceding the interim audit date and materially high after the interim date. This is because the error rates pertaining to the more recent period tend to have the greatest

impact on the ending balances (i.e., this is a justification in practice for the case 3 approach). Hence, the auditor takes pains to assure himself that there have been no major changes after his internal control tests at the interim date.

A related issue to the constancy of compliance error rates assumptions is the one about constancy of the dollar error generating process $F(\theta)$. However, in the present state of the art of auditing it appears auditors tend to view the system as if it were stable throughout the period. For example, audit textbooks (e.g., p. 203 Arens and Loebbecke) discuss system reliability in terms of the present even though the tests of transactions are made on a random basis covering the entire period. Hence it appears satisfactory to use stable error rates and error size functions $F(\theta)$ so that errors pertaining to ending balances are the same as those that held throughout the period.

If it can be agreed that the more complex system outlined above (i.e., figure 11) is more representative of an actual accounting environment the thing to consider now is whether the additional costs of creating such an environment are less than the benefits to be obtained from using it (compared to the simple system).

Before proceeding to the crux of the argument, it appears a proper perspective should be developed. It is obvious that there are certain aspects of the real world that are irrelevant no matter how completely one attempts to model the environment. For example, it is irrelevant to attempt to model physical mailing of confirmation when all that is needed in the simulation of an audit strategy is to provide a means for determining the difference between the book value and audit value of every randomly selected account balance. Similarly, the distribution of transaction values (which will be arbitrarily computed) is not relevant when the auditor wants to sample from the distribution of ending account balances (both book and audit values). This is because, concisely put, the relevant population for testing account balances is the accounts receivable file and not the transaction files. How far can this analogy be carried?

It appears to the researcher that the relevant attributes of an accounting environment that should be considered in the simulation are those pertinent to the goals of the simulated auditor (e.g., an accounts receivable file is necessary to simulate account balance testing). The audit strategies discussed in chapter four are a set of normative models which have been used in practice or proposed for meeting the particular goal of assessing the fair presentation of financial statement items through the use of statistical sampling techniques.

Let it be assumed now that all necessary assumptions have been made and an accounting system similar to that presented in figure 11 has been generated. Given that the goal of the auditor is to assess the fairness of the value represented by the sum of all the book values in the file, what kind of information will he be interested in obtaining? Without considering a particular audit strategy, the issue to resolve is what are the aspects of this environment considered relevant as covered in audit textbooks and in the professional literature. This appears a reasonable basis for deciding how the auditor would act.

First of all it goes almost without saying that the population of interest would be the accounts receivable file and that the sampling unit would either be a dollar or the individual accounts in the accounts receivable file. Assuming that this testing is done on a statistical basis, the auditor would like to know how much he could reduce such testing if he could rely on internal controls. There are several methods now available which tell the auditor how to do this and so what remains for him is to determine how much reliance to place on the internal control system for accounts receivable.

Assuming the auditor concludes the possibility of management over-override is remote, and that system design is adequate. (If either of these conditions did not hold the auditor would not attempt to rely on internal controls and reduce substantive tests; and therefore he would not test for compliance.) The auditor must identify the existing controls to prevent error and if he plans to rely on the pertinent controls, they must be tested through tests of compliance. (For example, see p. 177 of Arens and Loebbecke.)

Even if the complex system were used, it appears the simulation would have to be a closed system for the same reasons applicable to the simple system (i.e., because it is the upper bound on the value of internal control information that is of interest). Hence it would have to be assumed that the simulated auditor would be able to identify all pertinent controls and, furthermore, be able to determine what the material compliance error rate is for each control. (Even in the case of qualitative judgments about errors, the auditor must specify the kinds of errors he is fearful of and a critical error rate. In the extreme case of a very critical error, the auditor resorts to strict discovery sampling [in which case the hoped for error rate is zero]). Now, current trends in audit thinking appear to agree that the auditor must always, at least implicitly, consider the monetary impact of compliance error rates (e.g., see pp. 6-8 of Statistical Auditing by Don Roberts, or the Clarkeson, Gordon, & Co. manual p. 117, or even SAS No. 1 Sec. 320A.22). On reflection this is the most logical linkage to make.

In fact the internal control system structure of the simulation is based on this logic. Therefore, in testing the performance of an audit strategy, it is necessary to assume that the simulated auditor has sufficient expertise to be able to identify what a material attribute error rate is for each relevant attribute in the internal control system. (The simulation also attempts to model judgmental error but an important research question is: Given that the auditor does make accurate judgments can he rely on internal control information based on statistical tests of compliance to reduce his substantive tests?). The key point is that the auditor is able to somehow reach a decision about what error rates are critical for each of the internal controls (or internal control subsystem). He might have a model of the system in mind, he might be basing it on prior experience of the relationship of the attribute error rates to dollar errors in the accounts receivable file, he might have a general feel for these things as a result of professional experience and overall expertise. How he arrives at this judgment is not really germane to this study but the fact his judgment is consistent with the actual relationships of the simulated system is a key assumption in simulating the audit strategies.

By assuming that the auditor can accurately specify what the material error rates are for each attribute, it becomes clear why the complex system is no longer necessary to test the internal control hypothesis. Assume now that a complex system similar to figure 11 has been constructed including compliance error rates which result in an exactly material amount of dollar errors in the accounts receivable file. Be definition these attribute error rates are material and it is necessary to assume that the auditor has sufficient professional expertise to recognize this fact. Any time the auditor tests compliance he will be attempting to obtain sufficient assurance that compliance error rates are not equal to or above that considered material as defined by the system of controls for various files pertaining to the accounts receivable system. Essentially then the auditor should be relating compliance error estimates to the dollar accuracy of the records in the accounts receivable file.

How does he obtain these compliance error estimates? Well, basically, using sampling theory he computes a sample size for each attribute and then randomly samples all the populations of relevant attributes (relevant in the sense these attributes indicate the amount of dollar error in the accounts receivable file). The key aspect of all the attribute populations is that the sampling statistics are generally all based on the binomial distribution. That means that for a given error rate, once the population gets beyond a certain size, the same statistical results are obtained from a randomly drawn sample no matter how large the population. For example, essentially the same statistical properties hold for a random sample of 500 from a population with an attribute error rate of 5% regardless if the population consists of 7,000 or 70,000 or 700,000 items. In fact this is why most auditors use binomial tables based on an infinite population size (e.g., see p. 291 of Arens and Loebbecke). Of course, if the population gets sufficiently small and sampling is done without replacement, then the hypergeometric distribution must be resorted to, but all file sizes are sufficiently large in the simulation so that the binomial or Poisson provides good approximation. (The Poisson distribution is used as a conservative approximation to the binomial distribution which in turn conservatively approximates the hypergeometric which is the theoretically exact sampling distirbution of errors. See Harold J. Larson, Introduction to Probability Theory and Statistical Inference, (John Wiley & Sons Inc., 1969), pp. 118-119 and pp. 127-128.)

Recognizing this fact allows the introduction of considerable efficiencies in computational effort in simulating an audit strategy applied to the complex system. For example, instead of sampling randomly from the potentially hugh transaction file directly, it would be far cheaper to sample from a considerably smaller file having the same error rate. Note the statistical results for a given sample size would be the same, the only difference is that the sampling would not exactly physically mimic actual random selection from the much larger transaction file. But how important is it to get such mimicking in the simulation when the statistical results would not be changed?

This reasoning can be carried even further to support use of the simple system in the simulation. Assume that in the complex system certain relationships between compliance deviations and dollar error rates have a particular impact on the dollar accuracy of the accounts receivable file. It has just been argued that once the complex system has been set up there is no statistical need to sample directly from the full transaction file when sampling from a much smaller representative subset is just as valid for a particular sample size. This realization inevitably raises the question: Why not let the smaller attributes file define the same amount of error that the original larger file did? That is, given that one constructs a complex system with compliance error rates that result in a given total amount of dollar error, is it not true that statistically speaking an audit strategy would perform exactly the same way if a smaller attributes file were used which resulted in the same dollar error in the accounts receivable file? The answer is yes because the same statistical information is obtained by the audit strategy. Therefore, in terms of impact on audit strategy, any complex system is reducible to a comparable simple system. Since use of the simple system not only avoids the increased cost of sampling from larger (transaction) files but also the increased costs of generating the more complex system itself, it appears that the simple system is generally to be preferred over the complex one because the results are at least as generalizable. (At least generalizable because use of the complex system requires much more exact specification of the accounting environment [e.g., distribution of transaction book and audit values] with attendant assumptions which are open to question.)

The preceding discussion related primarily to a case 2 assumption. However, there should not be any difficulties in seeing that the argument applies just as well to case 3 if one works with attributes as they relate to transactions. That is, the implications of a case 3 assumption is that one essentially gets a figure 2 setup but the transaction files are reduced in number and size, generally speaking, to only those transactions pertaining to the outstanding ending balances. Again the auditor needs to relate attribute error rates (of relevant transactions only, now) to dollar accuracy of the accounts receivable population. And again, for statistical sampling purposes the simulation can be simplified by using a smaller attributes file with the same error rate and resulting in the same total dollar error (i.e., a simple system).

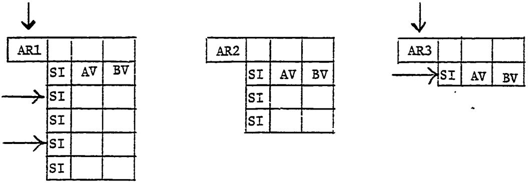
On the other hand, another way of justifying the simple system would be to redefine the attributes to relate them more directly to the level of aggregation of interest (in this case the population of accounts receivable), and then specify how these aggregated compliance errors affect the dollar accuracy of the accounts. For example, a particular compliance deviation in any of the unpaid sales invoices at year end could conceivably affect the dollar accuracy of the related outstanding balance because it might affect the accuracy of the recording of the transaction. However, the attribute now arises if any one condition in a set of attributes (defined by the set of relevant transactions) occurs. This approach really requires the ability to group transactions by account balances as illustrated in figure 10 and it appears possible for auditors to do this because this is the approach effectively followed by auditors when alternative procedures are resorted to in the case of non-responses to positive conformations. (see pp. 332-334 of Arens and Loebbecke.)

Current trends in auditing research appear to be recognizing the fact that one must distinguish that the impact of particular compliance deviations may be different depending on the level of aggregation of the population of interest. For example, both Cushing and Neter and Yu define reliability as the <u>probability</u> of correct processing. The key point is that one must be careful about specifying what kind of reliability one is referring to, e.g., the probability that the system correctly processes a sales invoice or the probability that the system correctly processes an account balance. Assuming by correct processing is meant that no monetary errors occur, the following figure illustrates the two probabilities are not necessarily equal.

The horizontal errors indicate monetary errors in the unpaid sales invoices for each account receivable in the file (this file has only three outstanding account receivables at year end with a supporting

FIG. 12

SIMPLE VS. AGGREGATE ATTRIBUTES



file of a total of ten unpaid sales invoices). AR1, AR2, and AR3 are used to designate the individual accounts. The vertical arrows indicate those accounts which have a monetary error. By definition a monetary error in a sales invoice (identified by an SI) results in a monetary error in the related account. The key aspect to note about this system is that although the monetary error rate for invoices is 30% (three invoices in error out of a total population of ten) the monetary error rate for accounts receivable is almost 67% (two accounts in error out of three)--although in a large files such an extreme difference is unlikely.

The upshot of this example is that in discussing the reliability of an internal control system, it may be important to specify what level of aggregation is being represented. In testing accounts receivable the direct probability of interest is really the probability of "correct" processing of an account (not the invoice, although knowledge of reliability of invoice processing is useful for determining reliability of account processing) and, therefore, attributes should be redefined to reflect this higher aggregation level. In fact this is required if one wants to implement a completely integrative model such as Grimlund's most general model for use with an accounts receivable file. (See pp. 90-97 of Grimlund, "A Framework for the Integration of Auditing Evidence" for assumptions that may need to be made in aggregating transaction data to effects on account balances).

Thus the level of aggregation represented is essentially a function of how the attributes are defined. For example, the more aggregated outstanding account balances, which are essentially the net result of a set of transactions, typically have more attributes (e.g., one attribute would be correct addition of the relevant component transactions--an attribute which does not apply to a particular transaction but does apply to the account balance). In addition the attributes tend to be more complex at the aggregate level because two or more control procedures might be necessary for a single purpose (e.g., a trade receivable resulting from two unpaid purchases requires that both purchases be properly processed in order that the receivable be processed correctly, or see figure 12 and the accompanying discussion). As stated in Sec. 320B.20 of SAS No. 1, "when two or more accounting control procedures are necessary for a single purpose, they should be regarded as a single procedure, and a deviation from any one of the sets should be regarded as a noncompliance occurence." On the other hand, Donald Roberts, on p. 7 of Statistical Auditing, states, "When two or more procedures are overlapping or duplicated to some degree the rate of noncompliance may correspond to lack of compliance with <u>all</u> procedures in the set." And it appears that a generalization of these two extremes is that an attribute can be defined as any combination of occurences of a set of attributes.

Thus a relevant set of attributes can always be defined for a particular level of aggregation represented by a record associated with the output of an internal control system (In fact Arens and Loebbecke p. 301 consider that "The definition of the attribute is a critical part of the use of attributes sampling."). This is also indicated in considering the issues involved in defining an attribute. (For example, see Roberts p. 149.) Thus by redefining attributes and considering different attributes (e.g., summarization attributes), the simple system representation makes as much sense for accounts receivable as it does for representing line items. The only difference between the two files would be in the size of the values for each record and how the various attributes would be interpreted.

Note that although the reasoning just concluded supports the simple system representation, the interpretations attached to the reocrds can be essentially different from those given in the earlier simple system justification. That is, the earlier justification relies on the ability of the simply system to yield the same statistical results that the complex system yields (and the fact that the internal control information, no matter how complex the system, is reducible to a set of relevant error rate estimates); while the latter justification relies on viewing the accounting system somewhat differently (but not less validly) depending on the level of aggregation. Either case represents strong arguments for the simple system approach.

Finally, perhaps it ought to be re-emphasized that the degree of detail in a simulation should be a function of the goals of the research. For purposes of the proposed research, it is sufficient that the accounting environment provide all the data necessary to test the audit strategies of interest. Any additional data is essentially superfluous. The key question to answer then is this: Does the described accounting environment, the simply system, provide enough detail to answer questions about the performance of audit strategies in certain general situations (e.g., "good"internal controls vs. "bad")?

The researcher feels that this question can be answered affirmatively.

This appendix concludes with a quotation from a noted simulation text:

> As mentioned earlier, computer simulation becomes a legitimate research tool when known analytical methods cannot supply a solution to a problem. Once simulation is adopted, however, certain admonitions made earlier in regard to modeling must be re-examined. Principal among these considerations is the issue of detail.
>
> Our earlier remarks stressed the fact that the amount of detail in a model is generally inversely related to our ability to obtain an analytical solution. However, if the minimal model, which contains the least detail needed for useful study, is not amenable to analytical solution and we adopt a computer simulation approach, we have the prerogative of building as much detail into our model as we like without concerning ourselves about the absence of analytical solution. It is precisely

this descriptive ability that holds the most attraction for modelers, for the greater the amount of detail, the more realistic the model is and consequently the closer they expect the results of a simulation run to conform to reality. However, we find in practice that judicious restraint in regard to detail is often the better policy, for at least three reasons.

First, to include detail we must devote effort and time to preliminary observation of the individual characteristics of the system under scrutiny. This attention induces a cost, the justification for which a modeler has to weigh with respect to his objective. This issue arises in most simulations, and no universal course of action fits all contingencies.

A second cost of detail arises when programming the model for eventual running. Added detail requires added programming, especially for the contingencies that detail specifies. Moreover, the inclusion of great detail at the outset of a modeling effort makes the job of locating the sources of error in the resulting computer program especially difficult because of the many potential trouble spots.

Third, because of the increased number of functions that a computer program based on a detailed model must perform, we expect a concomitant increase in running time and, consequently, cost. Testing for special situations, along with the need to update and manipulate system attributes, contributes notable to this cost. This third reason for restraint is not always apparent until actual simulation begins and we see the computing cost quickly mounting.

These sobering remarks provide a useful perspective for evaluating the degree of detail worth having. Often investigators using computer simulation get a gross model running first and then introduce detail where this model provides inadequate answers. This bootstrapping approach makes for a more judicious allocation of a modeler's time and effort and reduces debugging and running times. When changes in a simulation model are anticipated, a modeler is advised to organize his simulation program so that minimal effort is required. The choice of computer language also plays a major role here.

George S. Fishman, Concepts and Methods in Discrete Event Digital Simulation, (John Wiley & Sons, 1973), pp. 18-19.

APPENDIX II

Establishment of Least Favorable Error Rate Relationship Assumption for a Series System

The purpose of this appendix is to show that making the assumption that all compliance or monetary error rates in a series system are equal results in a conservative assessment regarding relationship of several error rates. By conservative assumption is meant that which will result in the least likelihood of reliance on a set of internal controls as a result of compliance testing so that unwarranted reliance and hence the risk of Type II error is minimized. Unwarranted reliance increases the risk of Type II errors because the result of the reliance is to increase the planned β level for the substantive test, thus automatically increasing the actual β risk for the test (see chapter two for the conceptual relationship).

The interest in this problem arises because when more than one attribute is involved and the item of interest is the product of several estimates (i.e., the reliability of the entire series system), it is important to consider all the combinations of error rates which affect system reliability. Donald Roberts appears to have been the first auditor to explicitly recognize this problem on p. 147 of his Statistical Auditing, where he states:

> ...It may be possible that some combination of individual rates of compliance deviations could produce a potential material monetary error even though the rate for each set is below its

> threshold level. For example, some proportion of inaccurately prepared invoices together with some proportion of inaccurately recorded cash receipts could produce a material overstatement of accounts receivable. The risk of unwarranted reliance when the potential material error is caused by some combination may be well above the tolerable level.
>
> Can this disadvantage be overcome? Probably, but the technology required to do it is quite complex, and further development is required before any practical solution can be offered. In the meantime, the auditor needs to be aware that present practice does not give much information regarding the possibility that material monetary errors may arise from a combination of causes.

As a response to this challenge, the researcher proposes an approach consistent with the philosophy followed by DUS advocates. That is, to control the risk of unwarranted reliance (which can ultimately increase the risk of Type II error) by controlling this risk at the prespecified level for the worst possible situation that can arise.

Consider the internal control system decomposed into n independent controls or subsystems (e.g., key controls, see page 122), each with an error rate (either compliance or the equivalent monetary one) p_i for control i. Let $\bar{p} = \frac{\sum_{i=1}^{n} p_i}{n}$ be the average of the p_i's. Now, for any given value of $\bar{p}$ the reliability of the series system is

$$R = \prod_{1=1}^{n} (1-p_i) = (1-p_1)(1-p_2)\ldots(1-p_n).$$

That is, R is the probability that the system processes a document without an error (where error can be defined to be some aggregate compliance error or a monetary error). Similarly, the error rate for the system (the probability that a record will have an error associated with it) is

$$1 - R = 1 - \prod_{i=1}^{n} (1-p_i).$$

With this notation the question now can be rephrased to be: What is the least favorable assumption to make about the system error rate 1-R given that the average of the component error rates p_i is some constant $\bar{p}$? That is, the auditor wants to make sure that the component error rates do not interact so that they produce a material aggregate system error rate $1-R = 1 - \prod_{i=1}^{n} (1-p_i)$. How can he obtain such assurance? The most conservative assumption, obviously, is to assume that the minimum possible system error rate for the possible combinations of p_i's results in a material system error rate. This is the most conservative assumption because it automatically rejects all other combinations of p_i that average to $\bar{p}$, including those that result in the maximum system error rate, whenever the minimum system error rate is rejected. That is, this assumption is the most likely to lead to a rejection of reliance on an internal control system represented by the possible error rates p_i which average to any constant amount $\bar{p}$. It is thus the least favorable error distribution assumption that can be made about the system for a given $\bar{p}$.

Given that the conservative assumption is to work with the minimum system error rate possible because it represents a worst case situation, this fact can not be used to determine what implications this conservatism has for the relationships of the error rates p_i. Now mathematics takes over entirely because the problem reduces to one of maximizing a function with a constraint.

Minimizing the system error rate 1-R is equivalent to maximizing system reliability $R = \prod_{i=1}^{n} (1-p_i)$. Thus the problem can be written as

maximize $\prod_{i=1}^{n} (1-p_i)$ subject to $\frac{\sum_{i=1}^{n} p_i}{n} = \bar{p}$ = constant and $0 \leq p_i \leq 1$

$i = 1,\ldots n$.

The solution to this problem was obtained by Teitlebaum and proves to be $p_1 = p_2 = \ldots = p_n = \bar{p}$, i.e., all the p_i's are equal to the constant $\bar{p}$.[1]

This result means that the least favorable assumption the auditor can make about the relationship of component error rates in a series system (and thus minimize the risk of unwarranted reliance) is to assume that they all equally contribute to produce an exactly material aggregate monetary error rate for the system.

The conservatism of this result makes intuitive sense because under this assumption the total material error rate is spread evenly a among the components so that the component errors rates are the smallest possible that are found to unacceptable.

It should be noted that equality of monetary error rates does not necessarily imply equality of compliance error rates in the real world. This is because there may be different relationships between monetary and compliance error rates among the different control points. However, in the simulation, this relationship is assumed constant

---

[1]See Teitlebaum, appendix II pp. 1-2 for the proof of this solution but in the context of a different problem.

(i.e., three to one) for all five internal control points or sub-systems. This is because such an assumption appears to be a conservative one for actual real world relationships. (See chapters four and five for more details.)

APPENDIX III

GLOSSARY

Alpha risk: α risk is the probability of making a Type I error. This may be an actual probability or a planned one. The planned probability is dependent upon the hypothesis or assumption made in planning the statistical test. Generally, the planned or nominal α risk is only an upper bound on the minimum actual α risk that is incurred. The dissertation obtains measures of the actual risk.

Audit risk: is the overall risk associated with applying the audit process or audit strategy. In its most general sense it includes non-sampling as well as sampling risks. However, in this dissertation the emphasis is on measuring the sampling risks because the major goal is to measure the statistical validity of the internal control hypothesis. Sampling risks take two forms: α risk and β risk (or combined risk when dealing with an audit strategy which incorporates the results of more than one statistical test).

Audit sampling strategy: has the same meaning as an audit statistical sampling strategy.

Audit statistical sampling strategy: is the abstraction of the audit process used in the dissertation. It consists of three stages of components which parallel the usual audit process: (1) obtaining a level of internal control information, (2) letting the information in stage (1) determine the extent of substantive testing via linkage rules, and (3) the substantive test. The goal of the audit process in this abstraction is assumed to be to reach a conclusion on the accuracy of the representation of the recorded amounts (book values).

Audit strategy: has the same meaning as audit statistical sampling strategy.

Attributes estimation: is the statistics of estimating rates and proportions of population attributes. This is the statistical theory that applies to compliance error rate, monetary error rate, and reliability estimation.

Beta risk: β risk is the probability of making a Type II error as a result of applying a single statistical test. This may be an actual probability or a planned one. Generally, if the test is statistically valid, the planned or nominal β risk is an upper bound on the actual β risk that is incurred. The dissertation obtains measures of the actual β risk.

Combined risk: is the probability of making a Type II error as a result of applying an audit sampling strategy. The combined risk for a particular substantive test equals the actual β risk associated with the test; however, the planned β risk can differ significantly from the planned combined risk and this difference is a function of the linkage rule and degree of reliance on internal controls. For example, planned combined risk in the simulation is always .05 but planned β risk can range from .05 to .5.

Compliance error rate: is the rate of compliance deviations from a particular accounting control procedure. This rate can be measured as a proportion of records having such deviations (estimated by random sampling without replacement) or the proportion of dollars associated with the records having such deviations (estimated by DUS for attributes). The former is implied in the text unless otherwise specified. Note that a compliance error rate can apply to either a single component or subsystem or as an aggregate output set of attributes to an entire system consisting of such components.

Compliance test: is a test "to provide reasonable assurance that the accounting control procedures are being applied as described" (SAS No. 1 Sec. 320.55). In this dissertation compliance tests are used to estimate the percentage of the sampling units that possess a particular characteristic (e.g., the proportion of K1 fields having a value of 1). Two types of sampling are simulated for this purpose: (1) statistical sampling of records without replacement and (2) dollar-unit sampling for attributes.

Dollar error: is the difference between the recorded value (book value) and the value that should have been recorded (audit value) where the difference is taken as follows: book value - audit value. Unless otherwise indicated by the text, dollar error is used to mean net total dollar error for the population.

Dollar error rate: is (total book value - total audit value)/(total book value)

Dollar-Unit sampling: DUS: DUS is a sampling selection technique wherein the probability of selection of a record is proportional to the recorded amount (book value) of the record. As in random sampling (where each record has an equal chance of selection regardless of the size of its recorded amount), the selection is conventionally done without replacement.

<u>Internal control hypothesis</u>: is the general hypothesis in auditing that substantive tests can be reduced as a result of reliance on internal controls without increasing the audit risks. This dissertation largely examines the purely statistical form of this hypothesis which is assumed here to be that: internal control information can be used to reduce the statistical sample size of substantive tests without increasing the actual audit risks that arise as a result of using an audit statistical sampling strategy.

<u>Monetary error rate</u>: is the proportion of either records or dollars having a dollar error. Unless otherwise indicated, the most frequent meaning in this dissertation is the proportion of records having a dollar error. It should be stressed that the monetary error rate does not necessarily equal dollar error rate, but that the amount of dollar error per sampled item does equal the amount of monetary error per sampled item. Also, a monetary error rate can apply to either a single component or subsystem, or to an entire system consisting of such components.

<u>Negative approach</u>: an approach used by auditors in constructing the statistical test in which the null hypothesis is that there is an exactly material amount of error in the recorded amounts. The chief result of this approach for substantive tests is that the statistical decision is based on estimated differences between the total audit

value and the recorded amount.

Null hypothesis: is the assumption which conventionally is used to control the critical error made in a statistical decision. However, under the positive approach in auditing this convention is violated and so the definitions of the Type I and Type II risks are usually the opposite of that in conventional statistical terminology. That is, in auditing the Type II error under the positive approach is the more serious error. In conventional statistical usage the Type I error is set up to be the more serious error.

Positive approach: is an approach used by auditors in constructing a statistical test in which the null hypothesis is that there are no errors (or, as indicated in appendix VII, that there are some immaterial errors amounting to M2) in the recorded amounts.

Reliability: is the probability of correct processing of a record in an internal control system or subsystem. In this dissertation, unless otherwise indicated, by correct processing is meant processing without a monetary error occurring in the record (or recorded dollar). Hence reliability = 1 - monetary error rate of system or subsystem, unless otherwise indicated.

Sampling risk: the risk associated with statistically testing only a samll proportion of a population and reaching a conclusion on the entire population.

Sampling Strategy: has the same meaning as audit statistical sampling strategy.

Strategy: has the same meaning as audit statistical sampling strategy.

<u>Substantive Test</u>: is a test of the numerical accuracy of the recorded amounts and hence usually involves variables estimation. Compliance tests on the other hand involves attribute estimation.

<u>Type I error</u>: is the error of incorrectly rejecting the accuracy of a materially correct book value (recorded amount) population. This convention has arisen in auditing due apparently to the pervasiveness of the positive approach in substantive tests. It should be noted this definition does not conform to statistical convention that the Type I error be the more serious audit error.

<u>Type II error</u>: is the error of incorrectly accepting the accuracy of a materially in error book value population. This is considered to be the more serious error in audit practice. (In fact according to Elliott and Rogers, "the minimization of the risk of Type II error is the reason for the existence of the auditing profession."...p. 49 of their paper.)

<u>Variables estimation</u>: is the statistical theory associated with estimating a quantitative characteristic such as a dollar amount of a population based on sample data.

<u>Unreliability</u>: is always 1 - reliability hence is dependent on the particular concept of reliability used. Since the most common concept of reliability used in this dissertation is 1 - system monetary error rate, the most frequent concept of unreliability is thus 1 - (1 - system monetary error rate) = 1 - 1 + system monetary error rate = system monetary error rate.

APPENDIX IV

Presentation of the Nancy R. Mann Method for the Computation of Lower Confidence Bounds on Series System Reliability[1]

The following notation is used in this appendix:

R_s = reliability of series system

K = number of independent subsystems

n_j = attribute sample size for subsystem j

x_j = number of failures or errors for subsystem j

$$1 - \hat{P} = \Pi_j \, [1 - (x_j/n_j)], \qquad \Delta = \frac{[\text{minimum of } (n_j - x_j)]^{-1} + \sum_{j=1}^{K} (n_j - x_j)^{-1}}{[\text{minimum of } n_j] \cdot [\sum_{j=1}^{K} n_j^{-1}]},$$

$$m_s = [\hat{P}/(1-.5\hat{P})] + (1/2)\Delta, \quad v_s = m_s(1/2)\Delta \quad \nu_s = 4m_s/\Delta$$

Then according to Mann, the lower bound on series system reliability with confidence level (c.l.) is computed as follows:

$$\text{Prob}\left\{ R_s \geq \exp[-m_s(1-(\frac{2}{9\nu_s}) + z_{c.l.}(\frac{2}{9\nu_s})^{1/2})^3] \right\} \doteq \text{c.l. for } \nu_s \geq 3$$

where $z_{c.l.}$ is the normal table value which includes an area of .5 - (1-c.l.).

---

[1] Nancy R. Mann, "Approximately Optimum Confidence Bounds on Series and Parallel-system Reliability for Systems with Binomial Subsystem Data," <u>IEEE Transactions on Reliability</u>, Vol. R-23, No. 5,

This bound is based on the approximation of the χ^2 distribution with ν_s degrees of freedom to the sample distribution of $2m_s(-\ln R_s)/\nu_s$. Thus, as discussed in chapter four, the bound is not exact but Mann calls it approximately optimal although its actual performance has been measured for relatively small (three or less) numbers of components only.

To obtain this bound one merely computes the numerical values indicated from the sample results, obtains the normal value $z_{c.l.}$ for the the confidence level desired and then computes the resultant value of the expression represented by exp [ ] above. This value is then the lower bound on system reliability with confidence level c.l.

It should be noted, again, that system failure x_j can be considered two different ways in the simulation. If it represents the compliance deviation (e.g., for use in linkage II(C)) then x_j/n_j is just the proportion of ones in the compliance test sample for attribute j. On the other hand, if by failure is meant a monetary error, then $(x_j/3)/n_j$ is used to estimate the number of equivalent monetary errors found in the compliance test sample for attribute j, (e.g., for use with linkage rule II(A)).

To compute the "C" value using the Mann formula, one must solve for the inverse of the formula given above. That is, it is necessary to solve for the confidence level associated with a given value of R (this is (1-significance level) of the given sample results for the null hypothesis of material errors: $R_s \leq .9$). Thus it is necessary to

---

December 1974.

now solve for $z_{c.l.}$ given $R_s = .9$ and compute the confidence level associated with the resultant $z_{c.l.}$ value. This is done as follows:

$$R_s = \exp[-m_s(1-(\frac{2}{9\nu}) + z_{c.l.}(\frac{2}{9\nu})^{1/2})^3] = .9,$$

implies $-m_s(1-(\frac{2}{9\nu}) + z_{c.l.}(\frac{2}{9\nu})^{1/2})^3 = \ln .9,$

implies $(1-(\frac{2}{9\nu}) + z_{c.l.}(\frac{2}{9\nu})^{1/2})^3 = \ln(.9)/(-m_s),$

implies $1-\frac{2}{9\nu} - z_{c.l.}(\frac{2}{9\nu})^{1/2} = [\ln(.9)/(-m_s)]^{1/3},$

implies $z_{c.l.} = (\frac{2}{9\nu})^{-1/2} [(\frac{\ln(.9)}{-m_s})^{1/3} + \frac{2}{9\nu} - 1)].$

Since all the values on the right hand side of the equality are known from the compliance test results, a value for $z_{c.l.}$ is obtained and this in turn allows computation of the associated confidence level c.l. (using the reverse of the procedure in looking up the $z_{c.l.}$ value for a given confidence level). This confidence level is now presumed to be the auditor's estimate of the "C" value to use in linkage II(B).

As discussed in chapter four, because of the conservatism of the crude method described there, the Mann method is used for all calculations involving system reliability. Unfortunately, the Mann method proves to be somewhat optimistic as is shown in chapter five, and this optimism is particularly serious when using linkage rule II(B). This optimism increases the risk of unwarranted reliance and hence the risk of Type II errors. Thus it is necessary to make adjustments to the

Mann method when using linkage II(B) to make it more reasonable for use in practical auditing. However, the problems associated with this optimism are less compared to those posed by the gross conservatism of the crude method, and so the Mann method is preferred for use in the simulation of the audit sampling strategies.

The adjustment to the Mann method when using linkage II(B) is described in chapter five.

APPENDIX V

Introduction to the Felix-Grimlund Model Formulas

As discussed in the Felix-Grimlund paper, they propose a Bayesian model based on what they call a beta weighting of normal distributions.[1] That is, they assume the posterior distribution of monetary error rates with sample size n and k errors can be represented by a standardized beta distribution $f_\beta(\rho/k'' = k'+k,\ n''=n'+n)$, and that the individual error sizes are realizations from a normal distribution process $f_N(\Pi)$ with unknown mean μ and unknown variance σ^2. The posterior distribution for the total amount of dollar errors is argued to be $f_{\beta N}(\Pi_T) = \int_0^1 f_\beta(\rho|k'',n')\ f_N(\Pi_T|a\rho,\frac{1}{b\rho})d\rho$ for $\nu'' > 2$

where $a = \chi m''$, $b = \chi[(\frac{\nu''}{\nu''-2})(\frac{1+k''}{k''})\ v'']$,

and $m'' = \frac{k'm'+km}{k'+k}$, $\qquad m = \frac{1}{k}\sum_{i-1}^{k} x_i$,

$\nu'' = \nu' + \nu + 1$ for $k' > 0$,

$\nu'' = \nu' + \nu$ for $k' = 0$,

$k'' = k' + k$, $v'' = \frac{[\nu'v' + k'm'^2] + [v\nu + km^2] - k''m''^2}{\nu''}$

---

[1]William L. Felix Jr. and Richard A. Grimlund, "A Sampling Model for Audit Tests of Composite Accounts," Journal of Accounting Research, Spring 1977, pp. 23-41.

$$\text{var}(\Pi'') = \frac{\nu''}{\nu''-2}\left(\frac{1+k''}{k''}\right)v'', \quad v = \frac{1}{k-1}\sum_{i=1}^{k}(x_i - m)^2 \text{ and } v=0 \text{ if } k = 1,$$

and $\nu = k - 1$, where x_i, $i = 1, k$ are the observed error amounts from a sample of size n.

Π_T = total dollar error amount.

Note that the requirement $\nu'' > 2$ implies that $K'' > 3$ which is not necessarily achieved for all environments for the sample sizes used in the simulation. Thus the following rules are used in the simulation.

Strategy 15: No internal control information:

Let $\nu' = 0$, $K' = 0$, $v' = 0$, then $\nu'' = \nu = K-1$, $K'' = K$, $v'' = v$.

$m'' = m$, and $\text{var}(\Pi'') = \frac{\nu}{\nu-2}\left(\frac{1+K}{K}\right)v$ by the above formulas.

(1) If $K > 3$, use above formulas directly to compute $f_{\beta N}$.

(2) If $2 \leq K \leq 3$, let $m'' = m$ and $\text{var}(\Pi'') = v$.[2]

(3) If $K < 2$, automatically accept the population.[3]

Strategy 6: Omniscient Case

As discussed in chapter four (pp. 227-8), $n^* = 750$ for the omniscient case thus for E1

$$K' = \rho' (750) = (.01)(750) = 7.5$$

Now, $\nu' = K' - 1$, hence $\nu'' = K'' - 1$ since $K' > 0$

For the preposterior analysis assume hypothetical sample results are the following:

---

[2]This rule is suggested by Felix-Grimlund on p. 38.

[3]Given that the dollar-unit sample would be highly likely to accept this result under the simulated environmental conditions, it was decided to use this as an acceptable decision rule for the Felix-Grimlund model.

$k = (\frac{k'}{n'})n$, $m = m'$, $\nu = k - 1$

$v = var(\Pi) = (\frac{\nu'}{\nu'-2})(\frac{1+k'}{k'})\ v'$ for hypothetical sample size n, where m', v', and ρ' are the parameter values used in the simulation (i.e., $m' = .5$ $v' = .057692$ and ρ' depends on the monetary error rate of the particular environment). With this information it is possible to compute the hypothetical posterior distribution $f_{\beta N}$ from m'', ν'', k'', and v'' using the formulas above; and then use this hypothetical posterior in Step 2 described on p.207-8 in chapter four.

Once the sample size n for substantive tests is so determined, an actual sample of n is taken and, then, the actual substantive test results for k, m, v, ν, and n are used to obtain the posterior values m'', ν'', k'', v'', and var (Π'') which are in turn used to compute the probability associated with the material amount of error.

Strategy 12: Objective Case

Now, $n^* = 250$ so that $k' = \tilde{\rho}\cdot(250)$ after compliance testing as explained on pp. 203 of chapter four. Also, as discussed in chapter four (p. 204) it is assumed the auditor can correctly specify the mean and variance of the process used to build the simulated audit environments. Thus $m' = .5$ and $v' = .057692$.

$$\text{Let } \tilde{\rho} = 1 - (1 - \frac{y_1}{450}) \cdot (1 - \frac{y_2}{450}) \cdot (1 - \frac{y_3}{450}) \cdot (1 - \frac{y_4}{450}) \cdot (1 - \frac{y_5}{450})$$

be the system monetary error rate estimate which is based on a compliance test sample size of 150 for each control point and assuming there is a 1/3 probability that each detected compliance deviation results in monetary error. If y_i is the number of compliance deviations found for internal control point i, then $\frac{y_i}{150}$ is the maximum likelihood estimate of the attributes error rate and 1/3 this number (i.e., $1/3 \cdot \frac{y_i}{150} = \frac{y_i}{450}$) is the estimate of the monetary error rate generated by the internal control point (or the rate at which particular monetary errors pass through the internal control point undetected).

Then $k' = \tilde{p} \cdot 250$, $\nu' = k' - 1$ and $\text{var}(\Pi') = (\frac{\nu'}{\nu'-2})(\frac{1+k'}{k'})\, v'$, whenever $\nu' \geq 2$; otherwise, planned substantive test sample size is automatically increased by ten in the preposterior analysis.[4]

After obtaining the substantive test sample evidence m, v, k, and actual n, compute $m'' = \frac{k'm'+km}{k'+k}$, $\nu'' = \nu' + \nu + 1$, $n'' = n + n'$

$$k'' = k' + k, \quad v'' = \frac{[\nu' v' + k'm'^2] + [\nu v + km^2] - k''m''^2}{\nu''},$$

$E(\Pi) = m''$, $\text{var}(\Pi'') = (\frac{\nu''}{\nu''+2})(\frac{1+k''}{k''})v''$ if $\nu'' > 2$,

$\text{Var}(\Pi'') = v''$ if $1 \leq \nu \leq 2$, and automatic acceptance of the population if $\nu'' \leq 1$.[5]

Once a posterior or preposterior $f_{\beta N}(\Pi_T)$ distribution with parameters k'', v'', m'', $\text{Var}(\Pi'')$, n'', ν'' has been computed for the total error amount, it is computationally useful to use an approximating distribution for which closed form analytical expressions are available. Although Felix-Grimlund suggested using a three parameter gamma

---

[4] This rule will automatically result in paradoxically large planned substantive test sample sizes when the internal control system is very reliable. Nevertheless, it is felt necessary to use this rule to illustrate the deficiencies of the present model.

[5] Again, since under these conditions the DUS TACS evaluation

distribution approximation in their article, more recent communication with Grimlund resulted in the suggestion that a four parameter extended beta distribution approximation is even more accurate. Hence, the extended beta approximation to the beta-normal is used in the simulation.

The well known statistical method of moments is used to approximate the beta-normal by the extended beta distribution. Basically, this means that once the parameter values of the beta-normal are computed the associated first four moment values of this distribution can be computed, and these then are assumed to be the first four moment values of the approximating extended beta. From the four moments of the approximating extended beta, the four parameter values of the extended beta can be calculated and then, via a transformation, the extended beta is converted to a standardized beta. It is necessary to convert to the standardized beta because this is the distribution for which cumulative probabilites have been tabulated. Once the beta-normal model has been approximated by a standardized distribution, it is a straightforward matter to compute the probability associated with any given amount of total dollar error. This then is used to compute the probability associated with the material amount of error (.05 x total book value) and the population is rejected (or substantive test sample size is increased) if this probability (which is

---

procedure would result in acceptance of the sample with a very high probability, this appears to be the most logical rule to use.

the combined risk level according to Felix-Grimlund) is greater than .05.

Thus, now it is necessary to give the four moment expressions for the beta-normal, the extended beta, and the standardized beta. These are all developed in Grimlund's dissertation. Grimlund has shown shown that the first <u>non-central</u> moment of $f_{\beta N}(\Pi_T)$ is $\mu_1'(\beta N) = a\mu_1'(\beta)$, where $a = \chi m$ and $\mu_1'(\beta)$ is the first <u>non-central</u> moment (i.e., the mean) of the standardized beta component of the beta-normal distribution.[6] The second <u>central</u> moment of $f_{\beta N}(\Pi_T) = \mu_2(\beta N) = b\mu_1'(\beta) + a^2\mu_2(\beta)$, where $b = \chi(\frac{\nu}{\nu-2})(\frac{1+k}{k})$ v and $\mu_2(\beta)$ is the second central moment (i.e., the variance) of the standardized beta component of the beta-normal distribution.

The third central moment of $f_{\beta N}(\Pi_T) = \mu_3(\beta N) = 3ab\mu_2(\beta) + a^3\mu_3(\beta)$ where $\mu_3(\beta)$ is the third central moment of the standardized beta component of $f_{\beta N}(\Pi_T)$.[7]

The fourth central moment of $f_{\beta N}(\Pi_T) = \mu_4(\beta N) = 3b^2[\mu_2(\beta) + \mu_1'^2(\beta)] + ba^2b[\mu_3(\beta) + \mu_1'(\beta)\,\mu_2(\beta)] + a^4\mu_4(\beta)$ where $\mu_4(\beta)$ is the fourth central moment of the standardized beta component of $f_{\beta N}(\Pi_T)$.

[6] Note that there are two beta's involved in the theory here. The first beta is used to model the uncertainty about the error rate, the second beta is used to approximate the distribution for total error $f_N(\beta_T)$ of which the first beta f_β is a component. The empirical results indicate that with DUS the first two parameters (p,q) of the approximating beta tend to converge to the two parameters of f_β as the sample population gets large. See Grimlund (pp. 218-220) for properties of the convergence of the skewness and kurtosis.

Also, there is an error in Grimlund's exposition at this point because he fails to distinguish between central and non-central moment monents in his formulas, p. 216 of Grimlund.

[7] Ibid., pp. 216-217.

To compute the above four moments of $f_{\beta N}(\Pi_T)$, one needs to know how to compute the associated moments for the standardized beta distribution with parameters k and n.

$$\mu_1'(\beta) = \frac{k}{n}, \quad \mu_2(\beta) = \frac{k(n-k)}{n^2(n+1)}, \quad \mu_3'(\beta) = \frac{2k(n-k)(n-2k)}{n^3(n+1)(n+2)},$$

$$\mu_4(\beta) = \frac{3k(n-k)[2n^2+k(n-k)(n-6)]}{n^4(n+1)(n+2)(n+3)}. \quad [8]$$

Note that in preposterior analysis $n = n^*$ and $k = \tilde{p} \cdot n^*$ in the above formulas. Thus, with the formulas given so far, it is possible to compute the four moments of the distribution of total dollar error Π_T from the sample results of either just compliance testing (using the rules given earlier) or the final posterior distribution after substantive testing.

Anyway, using the method of moments these four moments are assumed to equal the associated moments of the approximating extended beta distribution.

The extended beta distribution is really a generalization of the standardized beta distribution in the sense that whereas the latter is defined over the unit interval [0, 1], the former is defined over any closed interval [a,b]. Thus the extneded beta has the following form:

$$f_\beta(t|p,q,a,b) = \frac{(t-a)^{p-1}(b-t)^{q-1}}{(p,q)(b-a)^{p+q-1}} \qquad a \leq t \leq b, \quad p > 0, \quad q > 0,$$

$n = p + q > 0$. (Note the standardized beta distribution function has the same form except b = 1 and a = 0, i.e., they are constants. Also,

---

[8] Ibid., p. 113.

note that using the previous notation for the beta, p = k and q = n-k.)

Grimlund has shown how knowing the formulas for the four moments of this extended beta distribution it is possible to compute the parameters of this distribution p, q, a, b. The results of his analysis are given here (with some minor corrections to his original dissertation formulas given by *). Now, at this stage of analysis the four moments of the beta-normal $\mu'_{\beta N}(1)$, $\mu_{\beta N}(2)$, $\mu_{\beta N}(3)$, $\mu_{\beta N}(4)$ are assumed known and set equal to the four moments of the approximating extended beta distribution μ'_1, μ_2, μ_3, and μ_4. Hence the four moments of the approximating extended beta are assumed known and the goal is to compute the four parameters p, q, a, and b of this distribution.[9] Grimlund shows how to do this based on a technique by Elderton and Johnson.

Let $\beta_1 = \mu_3^2 / \mu_2^2$ and $\beta_2 = \mu_4 / \mu_2^2$

Then compute $n = \dfrac{6(\beta_2 - \beta_1 - 1)}{3\beta_1 - 2\beta_2 + 6}$, *

then $E = \dfrac{n^2}{4 + (1/4)\beta_1 \frac{(n+2)^2}{n+1}}$,

then $I = n\left(\dfrac{\mu_2(n+1)}{E}\right)^{1/2}$;

---

[9] There are other more simple approaches that can be followed in obtaining these parameter values. One of which is to predict the maximum of error in each direction, net overstatement and net understatement, consistent with what DUS practitioners do in practice. How However, Grimlund feels that a four parameter approximation is more accurate and so this is the method used in the simulation.

now, $p = 1/2[n \pm (n^2 - 4E)^{1/2}]$, and $q = n-p$

where the appropriate root is such that if $\mu_3 > 0$ then $p<q$, and $p > 0$ and $q > 0$.

Then, $a = \mu_1' - I(p/n)$ and $b = I+a.$* [10]

Thus all the parameters of the extended beta can be computed and so now the total error distribution is converted to the known extended beta distribution.

This in turn allows the ready computation of the probability of a given amount of error based on the auditor's posterior distribution reflecting all information available to him. This probability can be computed by using an inverse function procedure either for the extended beta distribution directly or by first transforming the extended beta to the equivalent standardized beta. The simulation used the transformation to the standardized beta distribution because of the availability of an integration procedure for this distribution.

Thus after obtaining the parameters p, q, a, b for the approximating extended beta, the following transformation was used. Let the standardized beta have the parameters p and q respectively from the extended beta, and the proportion of the domain for which the cumulative probability is computed is given by $X = \frac{M - a}{b - a}$, where M = the material amount = .05X total book value = 683575.15. Then the probability of material errors is the evaluation of the integral of the standardized beta function from 0 to X.

---

[10]Grimlund, pp. 169-170.

APPENDIX VI

Statistical Decision Rule Under the Negative Approach Using Stratified Mean-Per-Unit Estimation

This is the rule used in all stratified mean-per-unit tests unless otherwise specified. Using the notation given on pp. 217 , the one sided hypothesis test for overstatements is based on the following relationships:

$$E = Y - X \leq Y - (\hat{X} - A')$$

where Y = total book value (which is assumed known)

X = total audit value

$\hat{X}$ = estimate of total audit value

A' = achieved precision which now is equal to

$z_\beta \sqrt{\Sigma N_i(N_i - n_i)\frac{S_i^2}{n_i}}$ (compare to formula on p.216, i.e., the confidence coefficient and hence confidence level of the test is $1 - \beta$ as opposed to $1 - \alpha$ under the positive approach).

E = total net amount of overstatement.

The decision rule is thus: if $Y - (\hat{X} - A') \geq M$, then reject the total book value; otherwise, accept it.

Note that the positive approach described on pp.217 is statistically the same as the negative approach described here. This is shown in chapter five. The things to note about this negative approach are that (1) it is much more evident that the decision rule is based on the amount of estimated total dollar error in the population; and (2) as sample size is reduced due to a larger β, the confidence coefficient Z_β automatically gets smaller--this relationship is necessary to preserve the validity of the statistical test as discussed in chapter five.

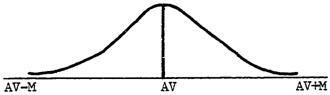
APPENDIX VII

CONTROL OF α RISK

The purpose of this appendix is to attempt to provide an intuitive explanation as to why the α risk can be expected to rise as the amount of dollar error in an accounting population increases. The basis for the illustration is the stratified mean-per-unit (STMPU) estimator which is used for several reasons: (1) it is probably the one most familiar to auditors in substantive testing, (2) its validity rests on the presumed normality of the sampling distribution of the estimator and the normal distribution is probably the most familiar one to illustrate the principles, (3) the high α risks associated with compliance testing and the Poisson approximation used in DUS have already been demonstrated in chapter four, and (4) the STMPU method is the one which explicitly attempts to control the α risk (although only for no errors as discussed on footnote 80 of chapter two).

Almost every auditor is familiar with the assumed normality of the distribution of the STMPU estimator of, say, the total audit value AV.[1] In illustrating the α risk, the following type of figure is frequently used.

Fig. 13: Normal Bell Shaped Curve of Sample Distribution of the Estimate $\hat{X}$ of the Total Audit Value



---

[1]By the central limit theorem of statistics, the sample average

In figure 13, the estimates are centered at the true AV (thus presuming there are no errors in the population since the maximum likelihood estimate of total error amount is unbiased) which is the mean of the above distribution; and there is some probability of obtaining estimates beyond a set amount AV+M, where M is here set equal to the materiality level (e.g., in the simulation M = .05 of the total book value). Since the simulation deals with one tailed tests only, the rest of this discussion will restrict itself to this situation. The principles remain the same as for the two tailed tests. Thus for any substantive test sample size n and assuming there are no errors the sample distribution looks as follows:

FIG. 14

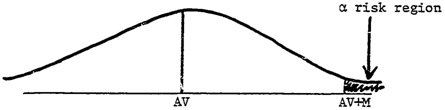


Figure 14 illustrates the α risk region associated with one tailed hypothesis test where the α risk is determined by the area under the curve to the right of AV+M. This region represents the probability associated with obtaining estimates beyond the AV+M value and hence the probability of rejecting the hypothesis that there is no error in the

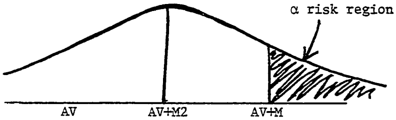
---

and sum of sample values will tend to be normally distributed. The larger the sample size the stronger the tendency.

accounting population. The size of this region is determined by the precision or variance of the distribution which, in turn, is determined by the sample size n for the substantive test. The larger the sample size, the smaller the variance and hence the smaller the α risk. Note that a rejection in this situation is incorrect and leads to a Type I error because the book value BV of the population is the actual audit value AV (around which the distribution is centered).

Now consider what happens to this rejection region as the amount of recorded error increases to say some arbitrary M2 value (i.e., BV = AV + M2) as illustrated in figure 15.

FIG. 15



Now the distribution is centered at AV+M2 and the whole distribution has shifted (due to the unbiasedness property associated with the mean estimator) toward AV+M by M2 thus increasing the region of rejection that there is no error. Note that this rejection is incorrect because the total amount of error is still immaterial. Thus the α risk has increased. <u>In general the probability of rejecting the hypothesis of zero errors had increased; and will always increase with increase in the amount of actual error in the population</u>. This explains why the α risk automatically increases with the amount of error.

Now, up to this point the error that has been discussed is the Type I error - the error of incorrectly rejecting a materially correct population, the probability of which is measured by the actual α risk. If the actual amount of error climbs up to the amount considered material M, or surpasses it, then rejection of the population suddenly becomes the correct decision and failure to reject, i.e., the complement of this probability of rejection, now becomes the probability of making an error. This kind of error is called a Type II error = the error of incorrectly accepting a materially in error population, and the associated probability is called the β risk (or combined risk, in the case of an audit strategy). Now the key thing to note is that as the actual population error gets bigger the probability of accpeting the population and hence β risk gets smaller. Therefore, the maximum β risk occurs at the exactly material amount of error. Thus control of β risk at a prespecified level is assured for any error amount BV-AV as long as the β risk is controlled at the exactly material amount of error BV-AV=M. This is the basis for controlling β risk in sample size planning.

Looking at it another way, the β risk at the exactly material amount is the probability of accepting the population; therefore, $1-\beta$ is the probability of rejecting (correctly) the population at the exactly material amount of error. However, consider what happens if the error amount shifted downward to just below the material level. Then the probability of rejection is just under $1-\beta$, but now this rejection is incorrect because by definition the population is acceptable (i.e., it does not have a material amount of error). Therefore, the maximum value the probability of incorrect rejection can can take must be 1-β. Again, this is because for any error larger than this, rejection no longer leads to an error. Thus even though the probability of rejection always grows with the amount of error, the probability of incorrect rejection is limited by M and the probability associated with correct rejection, which is 1-β. Thus the upper limit of the actual α risk is always 1 - actual β risk; and the α risk always grows from the actual α risk with zero errors to this maximum as the amount of error in the accounting system grows toward materiality.

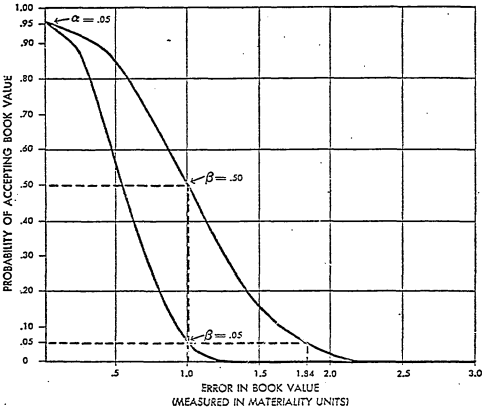
This phenomenon of the growth of the α risk does not appear to have been given sufficient recognition in the auditing literature. Possibly the best way to illustrate this growth in α risk is to use what are called power curves in statistics. This essentially has already been done in a prominent article by Elliott and Rogers which advocates the use of the hypothesis testing approach to statistical auditing and the control of both the α and β risks of the statistical test.[2] There graph is reproduced here as figure 16.

Figure 16 is based on the assumed normality of the distribution of sample estimator and using the hypothesis testing approach setting α = .05 and β ranging from .05 to .5 depending on the degree of reliance on

---

[2] This is the same Elliott and Rogers paper referred to earlier. See Robert K. Elliott and John R. Rogers, "Relating Statistical Sampling to Audit Objectives," Journal of Accountancy, July 1972, pp. 46-55. The graph is on their p. 51.

FIG. 16: ACCEPTANCE CHARACTERISTICS OF HIGHEST AND LOWEST RECOMMENDED β LEVELS



internal controls. The figure allows the calculation of the probabilities of accepting the hypothesis (vertical axis) that there are <u>no</u> errors in the population given that there is a certain amount as a proportion of what is considered material (horizontal axis).

Unfortunately, Elliott and Rogers did not discuss the implications of their diagram for the α risks associated with their parameter values although these are readily determinable from it. Perhaps this is because the β risk is the more serious risk associated with the substantive test. Nevertheless, by failing to point out the α risks associated with a substantive test for given levels of immaterial errors, it is very possible that many auditors may have been mislead about the extent of these risks. This is especially true considering that the planned β is the maximum such risk and thus without explicit warning many auditors may tend to believe that the maximum α risk is also equal to the planned $\alpha = .05$ level.

Since the simulation used the same planned α and β levels, it is instructive t compare the risks predicted by their graph and the risks mea d in the simulation.

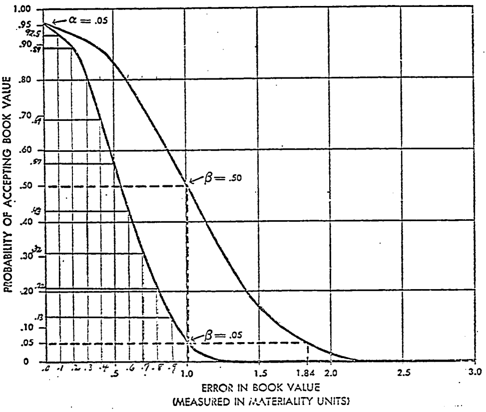
Firs of all note that as predicted earlier in this appendix, the probability of accepting the hypothesis of no errors in the book value drops as the amount of error increases. Equivalently, the probability of rejecting this hypothesis (the complement of the acceptance probability) climbs as the error amount decreases. Thus the probability of incorrectly rejecting is a maximum at the exactly material amount (i.e., the 1 value of the materiality axis because the proportion in terms of materiality is 1 at exact materiality) and is less at greater levels of error. This corroborates the earlier discussion here.

To help assess the α risk for various amounts of immaterial errors, figures 17 and 18 are provided. These figures compute the probabilities of acceptance of the zero error hypothesis for different levels of actual immaterial errors when the planned β is set to .05 and .5, respectively. The complement of this probability of acceptance is the probability of rejection which is the α risk when there is an immaterial error. Table 52 lists the α risks computed from figures 17 and 18.

Table 52 thus corroborates the earlier discussion in this appendix and indicates that statistical theory predicts the rise in α risk to the $1 - \beta$ level when the distribution of the sample estimate is normal. The fact that the simulation assessment of these risks is close to these predicted values (e.g., compare the actual α risks of table 15 and table 16 of chapter five to table 52 of this appendix), empirically supports the assumption of normality (or, more exactly, sufficient approximate normality) of the audit value estimator even for the very small sample sizes used in the study.

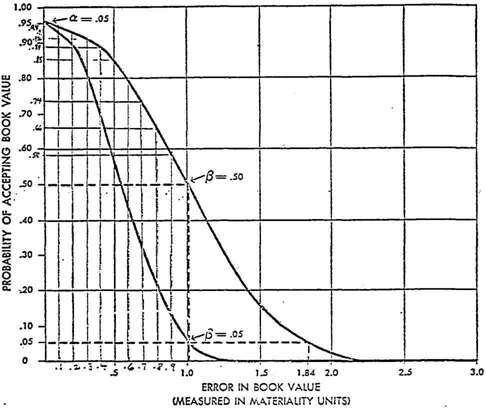
It may be of interest to estimate these characteristic curves for the substantive test methods used in the simulation. Figure 19 superimposes the estimated curve for DUS (dashed curve) and STMPU (dotted)

FIG. 17: ACCEPTANCE CHARACTERISTICS OF HIGHEST AND LOWEST RECOMMENDED β LEVELS



Computation of predicted acceptance probabilities for planned β = .05 and various immaterial error amounts.

FIG. 18: ACCEPTANCE CHARACTERISTICS OF HIGHEST AND LOWEST RECOMMENDED β LEVELS



Computation of predicted acceptance probabilities for planned β = .5 and various immaterial error amounts.

TABLE 52

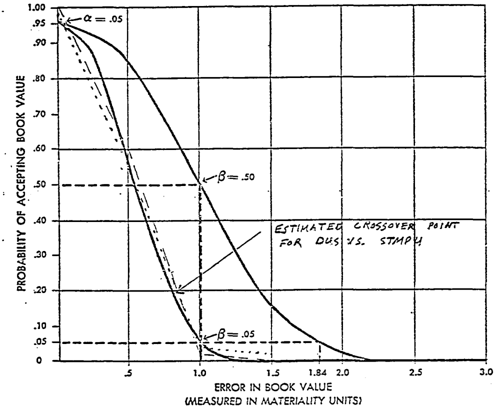
EXPECTED α RISKS ASSUMING NORMALITY OF ESTIMATOR

Amount of error in terms of Materiality M	Planned β = .05	Planned β = .5
.1M	.075	.06
.2M	.11	.07
.3M	.20	.08
.4M	.31	.11
.5M	.43	.15
.6M	.57	.20
.7M	.68	.26
.8M	.78	.34
.9M	.87	.42
.99M	.95	.50

curve for the planned β = .05 level (tables 13 and 15 respectively of chapter five). Figure 20 does this for the planned β = .5 level (tables 14 and 16 respectively of chapter five); and figure 21 indicates the characteristic curve for the Felix-Grimlund model at sample size of 120 (table 46 of chapter five).

These curves visually display the performance of these substantive test methods which are more completely discussed in chapter five. In particular figure 19 illustrates the location of the key crossover point of the probabilities which is at the heart of the controversy

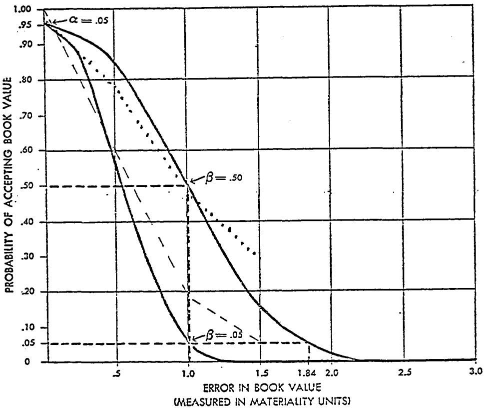
FIG. 19: ACCEPTANCE CHARACTERISTICS OF HIGHEST AND LOWEST RECOMMENDED β LEVELS



Dashed Line = Empirical probabilities for DUS with sample size of 120 and nominal β of .05.

Dotted Line = Empirical probabilities for STMPU with sample size of 237 and nominal β of .05.

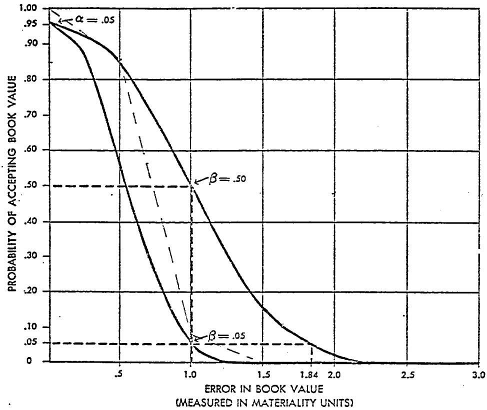
FIG. 20: ACCEPTANCE CHARACTERISTICS OF
HIGHEST AND LOWEST RECOMMENDED β LEVELS



Dashed Line = Empirical probabilities for DUS with sample size of 28 and nominal β of .5.

Dotted Line = Empirical probabilities for STMPU with sample size of 60 and nominal β of .5.

FIG. 21: ACCEPTANCE CHARACTERISTICS OF HIGHEST AND LOWEST RECOMMENDED β LEVELS



Dashed Line = Empirical probabilities for Felix-Grimlund test with sample size of 120.

revolving around the α risks associated with DUS and STMPU (See p.276-8 of chapter five for a discussion of this topic).

To help facilitate comparison of these performances, figure 22 is provided to indicate well performing substantive tests: DUS at sample size 237, and a theoretically perfect substantive test.

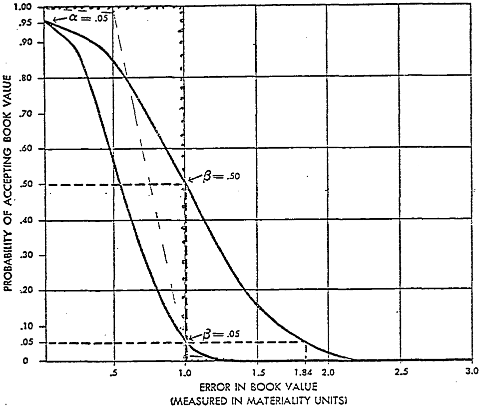
All the empirical probability curves are drawn by a straight line connection of the five actual empirical point observations. This results in some noncomparability of the theoretical with the empirical curves because not all turning points have been captured by the empirical data. However, comparisons between empirical curves are more valid. Overall the implications of these figures leads one to the same conclusions reached in chapter five.

Since the main purpose of this appendix is to point out the high α risks associated with substantive tests, particularly STMPU where such risks appear to be controlled for (although as it turns out this planned risk holds only for the unlikely case of no errors); it appears appropriate to indicate how to better control for such α risks. In particular, what may be needed in auditing is to derive a formula to control the α risk for a predefined amount of net immaterial errors, say M2, which is less than the exactly material amount M.

Kaplan has already developed a technique for doing this using DUS DUS; therefore, this appendix concludes with the comparable adjustment in STMPU sample size planning formulas.

This derivation closely parallels the logic used in deriving the formulas for controlling α risk assuming there are no errors in the

FIG. 22: ACCEPTANCE CHARACTERISTICS OF
HIGHEST AND LOWEST RECOMMENDED β LEVELS



Dashed Line = Empirical probabilities for DUS with sample size of 237 and nominal β of .05.

Dotted Line = A theoretically perfect substantive test which can always discriminate between material and immaterial total dollar error.

population. The researcher is indebted to Professor Robert B. Miller for showing the derivation of this earlier formula. The derivation of the formulas for control of the new α risk now follows.

Under the conventional positive approach the null hypothesis is that there are zero errors in the population so that Ho: BV = AV (using the notation of chapter four). A more realistic null hypothesis may be that Ho: BV = AV + M2 where M2 is some <u>immaterial</u> amount of error (assume net overstatement as in the simulation, hence M2 $\geq$ 0).

The alternative hypothesis is that there is an exactly material amount of error (to control β at its maximum level) thus H_1: BV = AV + M where M is the exactly material amount of total dollar error.

Under these conditions the auditor wants to compute a substantive sample size n such that the planned precision A controls the risk of Type I error at the level α. In formula terms:

$$\alpha = \text{Prob } \{ | \hat{X} - AV - M2 \ | \geq A \ | \ Ho: BV = AV + M2 \}$$

Similarly, the auditor wants the sample size to be such that planned precision A simultaneously controls the risk Type II error at the level β:

$$\beta = \text{Prob } \{ | \hat{X} - AV - M \ | \leq A \ | \ H_1: BV = AV + M \}$$

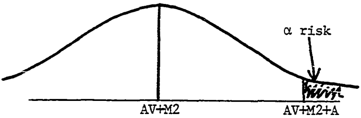
Both these formulas can be based on the assumed normality of the total audit estimate $\hat{X}$ which tends toward normality by the central limit theorem because it is based on the sum of the audit values found in the sample.

Now, by dividing both sides of the above inequalities by the sample estimate S_x of the standard deviation of the estimator, an approximate standard normal variable z results. Hence

$$\frac{\hat{X} - (AV + M2)}{S_x} \quad \text{and} \quad \frac{\hat{X} - (AV + M)}{S_x} \quad \text{are standard normal}$$

variates.

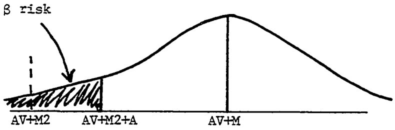
Thus under the null hypothesis, Ho: BV = AV + M2, the following situation arises:



That is, the mean of $\hat{X}$ is AV + M2, the assumed book value. To control α at the prespecified level, the precision A must be of such a value that the probability of getting an observation more than A distance from the mean of the estimates, AV + M2, is α or less. In terms of the standardized normal distribution this implies the standardized normal variate z must be such that

$$\alpha = \text{Prob } \{z \geq \frac{AV + M2 + A - (AV + M2)}{S_x} = \frac{A}{S_x}\}$$

Hence set $z_\alpha = \frac{A}{S_x}$ for controlling the α risk, where z_α is the standard normal value for .5 - α. Similarly, under hypothesis H_1: BV = AV + M the precision A is set so that the distribution looks as follows:



and so the auditor is interested in the negative (left) side tail of the standardizes normal distribution $\frac{AV+A+M2-(AV+M)}{S_x}$; so set $\frac{A + M2 - M}{S_x} = - z_\beta$ where z_β is the standard normal value for .5 - β. Thus the required precision A must satisfy both conditions simultaneously if the risks are to be controlled for at the planned levels. This means solving for both equations simultaneously: $\frac{z_\alpha}{-z_\beta} =$

$$\frac{\frac{A}{S_x}}{\frac{A+M2-M}{S_x}} = \frac{A}{A+M2-M}$$

, and this implies $z_\alpha \; (A+M2-M) = -z_\beta A$, and this implies that $A(z_\alpha + z_\beta) = -z_\alpha(M2-M) = z_\alpha(M-M2)$, and this implies

$$A = \frac{z_\alpha(M-M2)}{z_\alpha + z_\beta} \;. \qquad (1)$$

Note that if the amount of error for which the α risk is to be controlled for is zero (i.e., M2 = 0), then the planned precision reduces to the same formula as used in the simulation and in practice. (See p. 215 of chapter four.) Equation (1) is thus a generalization of the usual formula, where now M2 can be any error amount such that

$0 \leq M2 < M$.

The implications of this formula are relatively straightforward. If the auditor would like to control the α risk for an amount of error greater than zero, say to .5 of M, then the planned precision becomes half of what it is when α risk is only controlled at zero errors. Halving the planned precision means approximately quadrupling the plan planned sample size.[3] Thus sample sizes for STMPU may need to be much larger than previously suggested for controlling α risk at more tolerable levels.

This is precisely the same conclusion reached by Kaplan for DUS.[4] However, Kaplan appeared to imply that, as a consequence, sample sizes for DUS needed to be on a level for those with STMPU to control for the same levels of sampling risks. That this is not the case has been demonstrated in the study reported in this dissertation. Since at sample size 120 DUS has essentially the same sampling risks as STMPU at sample size of 237, an approximate quadrupling of both sample sizes should still result in half the STMPU sample size for DUS. In fact table 18 of chapter five indicates that even for a DUS sample size of

---

[3]Taking as an example the computations of p.266 of chapter five and letting planned precision be 1/2 x 341,787.5 = 170893.75 for α = .05 (but now for immaterial errors as much as .5M) and β = .05, the planned sample size then turns out to be

$$n = \frac{(1.65)^2(13671503)(729007.95)}{(170893.75)^2 + (1.65)^2(1438226100)} = 820 \text{ which is}$$

over 3.5 times the planned sample size under the old α risk definition (i.e., 820 vs. 225).

[4]Kaplan, "Sample Size Computations for Dollar-Unit Sampling," p. 131.

only 237 (i.e., less than 1/3 the planned STMPU sample of 820 is footnote 3), the actual α risk at .5M is already less than .05 thus further demonstrating the superiority of DUS for the simulated environments.

BIBLIOGRAPHY

American Institute of Certified Public Accounts, Statement on Auditing Standards, Codification of Auditing Standards and Procedures No. 1. New York: American Institute of Certified Public Accountants, 1973.

Anderson, Rodney J. The External Audit, Concept and Techniques. 2 vols. Toronto: Pittman Publishing Co., 1977. Vol. 1: The External Audit 1, Concept and Techniques.

Anderson, R.J., and Leslie, D.A. "Discussion of Considerations in Choosing Statistical Sampling Procedures in Auditing." Journal of Accounting Research 13 (Supplement 1975):53-64.

Arens, Alvin A., and Loebbecke, James K. Auditing, An Integrated Approach. Englewood Cliffs, New Jersey: Prentice Hall, Inc., 1976

Arkin, Herbert. Handbook of Sampling for Auditing and Accounting. Second Edition. New York: McGraw Hill Book Company, 1974.

Ashton, Robert. "An Empirical Study of Internal Control Judgment." Journal of Accounting Research 12 (Spring 1974):143-157.

Auditing Concepts Committee. "Report of the Committee on Basic Audit Concepts." Accounting Review 47 (Supplement 1972):57.

Bailey, Andrew D. Jr., and Jensen, Daniel L. "A Note on the Interface Between Compliance and Substantive Tests." Journal of Accounting Research 15 (Fall 1977):293-299.

Barlow, Richard E., and Proschan, Frank. Statistical Theory of Reliability and Life Testing Probability Models. New York: Hold, Rinehort and Winston, 1975.

Beck, Paul John. "A Critical Analysis of the Regression Estimator in Audit Sampling." Ph.D. dissertation, University of Texas, 1977.

Bodnar, George. "Reliability Modeling of Internal Control Systems." Accounting Review 50 (October 1975):753-768.

Brown, R. Gene. "Objective Internal Control Evaluation." Journal of Accountancy 126 (August 1968):49-56.

Buehler, Robert J. "Confidence Limits for the Product of Two Binomial Parameters." Journal of the American Statistical Association 52 (December 1957):482-493.

Burns, David C. "Computer Simulation: A Tool for Testing the Effectiveness of Internal Control Auditing Methods." Proceedings of the Fifth Annual Midwest AIDS Conference 1 (1974):D1-D4.

Burns, David C., and Loebbecke, James K. "Internal Control Evaluation: How the Computer Can Help." Journal of Accountancy 140 (August 1975):60-70.

Carmichael, D.R. "Behavioral Hypothesis of Internal Control." Accounting Review 45 (April 1970):235-245.

Clarkeson, Gordon, & Co. Audit Testing Course, Course Summary n.p.: Clarkeson, Gordon, & Co., 1975.

Cochran, William G. Sampling Techniques. 3rd ed. New York: John Wiley & Sons, 1977.

Corless, John. "Assessing Prior Distributions for Applying Bayesian Statistics in Auditing." Accounting Review 49 (January 1974): 556-566.

Cushing, Barry E. "A Mathematical Approach to the Analysis and Design of Internal Control Systems." Accounting Review 49 (January 1974):32-45.

________. "Decision-Theoretic Estimation Methods in Accounting and Auditing." Paper presented at the University of Wisconsin, December 2, 1977.

DeGroot, Morris H. Probability and Statistics. Reading, Massachusetts: Addison-Wesley Publishing Company, 1975.

Elliott, Robert K. and Rogers, John G. "Relating Statistical Sampling to Audit Objectives." Journal of Accountancy 134 (July 1972):46-55.

Ernst and Ernst. Audit Sampling. n.p.: Ernst and Ernst, 1976.

Felix, William L. Jr., and Goodfellow, James L. "Audit Tests for Internal Control Reliance." Paper presented at the American Accounting Association Convention, August 1978.

Felix, William L. Jr., and Grimlund, Richard. "A Sampling Model for Audit Tests of Composite Accounts." Journal of Accounting Research 15 (Spring 1977):23-40.

Fienberg, Stephen E.; Neter, John; and Leitch, R.A. "Estimating the Total Overstatement Error in Accounting Populations." Journal of the American Statistical Association 72 (June 1977): 295-302.

Fishman, George S. Concepts and Methods in Discrete Event Digital Simulation. n.p.: John Wiley & Sons, Inc., 1973.

Garstka, S.J. "Computing Upper Error Limits in Dollar-Unit Sampling." Journal of Accounting Research 15 (Autumn 1977):179-192.

Goodfellow, James L.; Loebbecke, James K.; Neter, John. "Some Perspectives on CAV Sampling Plans." CA Magazine, Part I--Oct. 1974, pp. 23-30; Part II--Nov. 1974, pp. 46-53.

Grimlund, Richard A. "A Framework for the Integration of Auditing Evidence." Ph.D. dissertation, University of Washington, 1977.

_______. "The Integration of Internal Control System and Account Balance Evidence." Working paper at the University of Iowa, August 1978.

Hamdi, Aly, and Duboff, Jack. "Statistical vs. Judgmental Sampling: An Empirical Study of Auditing the Accounts Receivable of a Small Retail Store." Accounting Review 46 (January 1971):119-128.

Haskins & Sells. Audit Sampling, A Programmed Instruction Course. n.p.: Haskins & Sells, 1970.

Hillier, Frederick S., and Lieberman, Gerald J. Operations Research. Second Edition. San Francisco: Holden Day Inc., 1974.

"Holding the Line on Audit Fees." Business Week, October 23, 1978, pp. 57-58.

Ijiri, Yuji, and Jaedicke, R.K. "Reliability and Objectivity of Accounting Measurements." Accounting Review 41 (July 1966): 474-483.

Ijiri, Yuji, and Kaplan, Robert S. "A Model for Integrating Sampling Objectives in Auditing." Journal of Accounting Research 9 (Spring 1971):73-78.

Joyce, Edward J. "Expert Judgment in Audit Program Planning." Journal of Accounting Research 14 (Supplement 1976):29-60.

Kaplan, Robert S. "A Stochastic Model for Auditing." Journal of Accounting Research 11 (Spring 1973):38-46.

_______. "Sample Size Computations for Dollar-Unit Sampling." Journal of Accounting Research 13 (Supplement 1975):126-133.

_______. "Statistical Sampling in Auditing with Auxiliary Information Estimators." Journal of Accounting Research 11 (Autumn 1973):239-258.

_______. "Synthesis." Journal of Accounting Research 13 (Supplement 1975):141-142.

Kinney, William R. Jr. "A Decision-Theory Approach to the Sampling Problem in Auditing." Journal of Accounting Research 13 (Spring 1975):117-132.

_______. "Decision Theory Aspects of Internal Control System Design/ Compliance and Substantive Tests." Journal of Accounting Research 13 (Supplement 1975):14-34.

Knoblett, J.A. "The Applicability of Bayesian Statistics in Auditing." Decision Sciences 1 (July-Oct. 1970):423-440.

Kotas, John F., Lau, Amy Hing Ling, and Lau, Hon-Shiang. "A General Approach to Stochastic Management Planning Models: An Overview." Accounting Review 53 (April 1978):389-401.

Larson, Harold J. Introduction to Probability Theory and Statistical Inference. n.p.: John Wiley & Sons, Inc., 1969.

Lloyd, David K., and Lipow, Myron. Reliability: Management Methods and Mathematics. 2nd ed. Redondo Beach, California: By the Authors, 1977. (1st ed. Englewood Cliffs, New Jersey: Prentice Hall, Inc., 1962).

Loebbecke, James K., and Neter, John. "Considerations in Choosing Sampling Procedures in Auditing." Journal of Accounting Research 13 (Supplement 1975):38-52.

Magee, Robert P. "Discussion of Auditors' Loss Functions Implicit in Consumption Investment Models." Journal of Accounting Research 13 (Supplement 1975):121-123.

Mann, Nancy R. "Approximately Optimum Confidence Bounds on Series and Parallel-system Reliability for Systems with Binomial Subsystem Data." IEEE Transactions on Reliability R-23 (December 1974):295-304.

Mann, Nancy R.; Schaefer, Ray E.' and Singpurwalla, Nazer D. Methods for Statistical Analysis of Reliability and Life Data. New York: John Wiley & Sons, 1974.

Mautz, R.K., and Mini, Donald. "Internal Control Evaluation and Audit Program Modification." Accounting Review 41 (April 1966):283-291.

Mautz, R.K., and Sharaf, Hussein A. The Philosophy of Auditing. Sarasota, Florida: American Accounting Association, 1961.

Mock, Theodore J. and Turner, Jerry L. "A Field Test of the Effect of Changes in Internal Controls in Audit Programs." Paper presented at the University of Wisconsin-Madison on April 5, 1978.

Neter, John. "Applicability of Statistical Sampling Techniques to the Confirmation of Accounts Receivable." Accounting Review 31 (January 1956):82-94.

Neter, John; Leitch, R.A.; Fienberg, Stephen E. "Dollar Unit Sampling: Multinomial Bounds for Total Overstatement and Understatement Errors." Accounting Review 53 (January 1978):77-93.

Neter, John and Loebbecke, James. Behavior of Major Statistical Estimators in Sampling Accounting Populations. Auditing Research Monograph 2. New York: AICPA, 1975.

Neter, John; Loebbecke, James K.; and Goodfellow, James L. "Some Perspectives on CAV Sampling Plans." CA Magazine, November 1974, pp. 25-26.

Ramage, J.G.; Krieger, A.M.; Spero, L.L. "An Empirical Study of Error Characteristics in Audit Populations." Paper presented at a University of Chicago symposium on empirical research in auditing, May 1979.

Reneau, J. Hal. "CAV Bounds in Dollar-Unit Sampling: Some Simulation Results." Accounting Review 53 (July 1978):669-680.

Rittenberg, Larry, and Schweiger, Bradley. "Auditor Reporting Responsibilities in a Changing Society." Wisconsin Working Paper 7-76-26, Graduate School of Business, University of Wisconsin-Madison, 1976.

Roberts, Donald M. "A Statistical Interpretation of SAP No. 54." Journal of Accountancy 137 (March 1974):47-53.

_______. "Discussion of The Real Risks in Audit Sampling." Journal of Accounting Research 13 (Supplement 1975):92-94.

_______. Statistical Auditing. New York: American Institute of Certified Public Accountants, 1978.

Robertson, Jack C. Auditing. Dallas: Business Publications, Inc., 1976.

Scott, William R. "A Bayesian Approach to Asset Valuation and Audit Size." Journal of Accounting Research 11 (Spring 1973):304-330.

Shooman, Martin L. Probabilistic Reliability: An Engineering Approach Approach. New York: McGraw Hill, 1968.

Sorensen, James E. "Bayesian Analysis in Auditing." Accounting Review 44 (July 1969):555-561.

Stratton, William O. "Accounting Internal Control Systems: Their Reliability and Dichotomic Structure Functions." Ph.D. dissertation, Claremont Graduate School, 1977.

Swieringa, Robert, and Carmichael, D.R. "A Positional Analysis of Internal Control." Journal of Accountancy 131 (February 1971): 34-43.

Teitlebaum, A.D. "Dollar-Unit Sampling in Auditing." Paper presented to the National Meeting of the American Statistical Association, December 1973.

Teitlebaum, A.D.; Leslie, D.A.; and Anderson, R.J. "An Analysis of Recent Commentary on Dollar-Unit Sampling in Auditing." Paper available on request from the authors.

Teitlebaum, A.D.; McCray, J.H.; Leslie, D.A. "Approaches to Evaluating Dollar-Unit Samples." Paper presented to the American Accounting Association Convention, August 1978.

Teitlebaum, A.D., Robinson, C.F. "The Real Risks in Audit Sampling." Journal of Accounting Research 13 (Supplement 1975): 73-74.

_______. "A Reply" Journal of Accounting Research 13 (Supplement 1975):95-97.

Toba, Yoshihide. "A General Theory of Evidence as the Conceptual Foun Foundation in Auditing Theory." Accounting Review 50 (January 1975):7-24.

Tracy, John A. "Bayesian Statistical Methods in Auditing." Accounting Review 44 (January 1969):90-98.

Vanecek, Michael T. "Bayesian Dollar-Unit Sampling in Auditing." Ph.D. dissertation, University of Texas, 1978.

Warren, Carl S. "Discussion of a Statistical Technique for Analytical Review." Journal of Accounting Research 13 (Supplement 1975):11.

Weber, Ronald. "Auditor Decision Making: A Study of Some Aspects of Accuracy and Consensus." Ph.D. dissertation, University of Minnesota, 1977.

Winkler, Robert. Introduction to Bayesian Inference and Decision. New York: Holt, Rinehart, and Winston Inc., 1972.

Winkler, Robert L., and Hays, William L. Statistics: Probability, Inference, and Decision. New York: Holt, Rinehart, and Winston, Inc., 1975.

Winterbottom, Alan. "Lower Confidence Limits for Series System Reliability from Binomial Subsystem Data." Journal of the American Statistical Association 69 (September 1974):782-788.

Yu, Seongjae, and Neter, John. "A Stochastic Model of the Internal Control System." Journal of Accounting Research 11 (Autumn 1973):273-295.

VITA

Waldemar John Smieliauskas, son of Edward and Irene Smieliauskas, was born May 12, 1947, in Rottenberg, Germany. His family immigrated to the United States in 1952 and settled in Chicago, Illinois. He graduated from Thomas Kelly High School, Chicago, Illinois, in June, 1965. He received his B.S. with honors in mathematics from the University of Illinois-Chicago Circle Campus in June 1970. He studied an additional year full time doing graduate work in mathematics before obtaining a position as a cost accountant in the evenings and obtained an MS degree in mathematics from the University of Illinois-Chicago Circle Campus in June 1973.

In August 1974 he left his position as regional cost accountant to enter the Ph.D. program in accounting at the University of Wisconsin-Madison. He obtained his CPA Certificate in February 1977 and is a member of the AICPA. He has been married to Chantal Houvenaegel since April, 1970, has one child, and currently is an Assistant Professor of Accounting at the University of Toronto, Canada.

TITLE OF THESIS SIMULATION ANALYSIS OF THE STATISTICAL VALIDITY OF THE INTERNAL CONTROL HYPOTHESIS OF AUDITING WITH IMPLICATIONS FOR SUBSTANTIVE TESTING METHODS AND LINKAGE RULES

Major Professor PAUL H. WALGENBACH

Major Department BUSINESS

Minor(s) MATHEMATICS

Full Name WALDEMAR JOHN SMIELIAUSKAS

Place and Date of Birth ROTTENBERG, GERMANY on MAY 12, 1947

Colleges and Universities: Years attended and degrees

UNIVERSITY OF ILLINOIS: 1965-1970, BS WITH HONORS IN 1970

UNIVERSITY OF ILLINOIS: 1970-1973, MS IN 1973

Membership in Learned or Honorary Societies NONE

Publications NONE

Date December 23, 1979