# Reliability Theory

# and

# Internal Control Evaluation

B.N.Srinidhi[1]

M.A. Vasarhelyi[2]

August 1984
**Revised April, 1985**

Support was provided by the Peat, Marwick, Mitchell Foundation as part of the Research Opportunities in Auditing Program. The views expressed here are our own and do not necessarily reflect the views of the Peat, Marwick, Mitchell Foundation. This support is gratefully acknowledged.

This working paper is in preliminary form and should not be quoted without explicit consent of the authors. Comments are requested. The comments and suggestions of Profs. D. Miller, G. Shillinglaw, J. Bell, W.Baber, S.El-Gazzar, J.Loebbecke, B.Cushing, Thomas Lin, Theodore J. Mock as well as at seminars at Columbia University, New York University, the University of Florida and the Conference on Audit Evidence at USC are also gratefully acknowledged.

[1] Assistant Professor, New York University

[2] Associate Professor, Columbia University

# 1 INTRODUCTION

Internal control systems are evaluated both as part of systems design and maintenance activity by managers and as part of evidence collection and inference activity by auditors. The necessity and importance of internal control evaluation has been recognized by the auditing profession for a long time.

The second standard of field work developed by AICPA[3] mandates an evaluation of the internal control system as a basis for restricting substantive tests. This is also reinforced by SAP (AICPA, 1979). SAS20 (AICPA,1977) [1] requires communication of internal control weaknesses to management. The design, compliance, and improved methods of evaluation of the Internal Control System have become important concerns of auditors.

This increased concern has resulted in a demand for structured mathematical modeling of the Internal Control Systems. ASOBAC [4] expresses this concern by stressing the need for inductive inference techniques in audit research. [Yu and Neter (1973) [23], Cushing(1974) [9] and Hamlen(1980) [15], Ishikawa(1975) [16], Stratton(1980) [21] and Srivastava(1981) [20]]. The auditing process consists of both a statistical inference process and an audit decision process. These models have all tried to improve the statistical inference process in internal control evaluation. However, there has been no attempt to (1) integrate the mathematical models with the decision process of the auditors, and (2) study their impact on the quality and nature of audit judgments. This may partly explain the apparent lack of interest in these models by the Auditing profession.

This study (1) distinguishes between the mathematical inference processes and the decision processes in internal control evaluation, (2) develops a normative quantitative internal control evaluation model which captures the mathematical relationships inherent in the process, (3) compares the normative model with the descriptive model of auditor judgments developed empirically, and (4) provides an empirical validation of the proposed quantitative decision aid.

Section 2 presents research objectives and reviews relevant issues. Section 3 provides a mathematical formulation of these decision stages to represent the relationships between inputs and outputs at each stage. Section 4 describes, presents the results and interprets an experiment to evaluate the influence of a reliability based decision aid. The final section summarizes key findings and presents conclusions.

# 2 LITERATURE REVIEW AND OBJECTIVE

While some behavioral studies have developed subjective criteria for decision evaluation, others, have empirically evaluated auditor judgments by applying subjective

---

[3] See AU150.02 - Standards of Field Work - AICPA

criteria to different audit situations. Little, however, has been done to incorporate normative models to capture the mathematical relationships and use them as aids to the decision process in auditing. Consequently, it has not been possible to transfer the insights gained in the behavioral studies to model development or vice-versa.

Such a transfer of insights is particularly important in view of progressive systematization of auditor judgments on internal controls. Cushing and Loebbecke [11] find that almost all the CPA firms have well defined procedures for recording and documenting internal control judgments.

Yu and Neter [23] treated error in accounting data as a stochastic variable and used probability transformation for each system element to trace the probability of error through the system. Cushing [9, 8] introduced reliability theory to evaluate internal control systems. Each transactions cycle was defined as an internal control system comprising of internal control "procedures". Different error types were recognized and each combination of error type and control procedure was represented by five probability parameters.

Bodnar [6] refined Cushing's model for human systems. Stratton [22, 21] represented complex internal control systems in a reliability network and simplified Cushing's five into a single parameter representation. However, such simplification was not supported by either analytical or by empirical studies. Srivastava [20] attempted to integrate sequence logic into reliability network representation. Grimlund [14] integrated the basic models of Cushing [9] and Yu and Neter [23] relating the evidence on internal control systems to judgments on the reported balances in financial statements.

Lens model studies [2] indicate that judgment quality can be evaluated by accuracy. However, where normative criteria are not available, judgment agreement measures (such as consensus and stability) have been proposed as evaluative measures. Einhorn [12] argues that consensus is a necessary condition of good judgment quality. Goldberg and Werts [13] identify stability, consensus and convergence as measures of quality.

Ashton [3, 2] evaluated the internal control judgment of auditors using consensus and stability measures. The internal control system was judged on a six point scale of its strength. A mean consensus of 0.70 and a mean stability of 0.81 were found. Joyce [17] used the same criteria for the same type of decision but the judgment was in terms of the influence of internal control strength on the extent of substantive testing. There are a number of similar descriptive studies which use these criteria in different audit situations. A listing of these can be found in Ashton [2]. Although there has been significant research interest in both descriptive and model building studies, an interface between the two is lacking.

This study provides such an interface. Simple and separable decision stages are identified, and the mathematical relationships in each stage are captured in probability-based models which can aid auditors to make decisions in each stage. A descriptive model of auditor decision process in these stages is developed from an empirical study and contrasted with the normative probabilistic model to demonstrate the need for using the latter as decision aid.

The influence of using[4] such a decision aid on the quality of their judgments is empirically examined. The criteria used in such an examination are those developed in behavioral decision making studies.

Next section identifies decision stages and develops the mathematical relationships in these stages of internal control evaluation.

## 3 ANALYSIS OF DECISION STAGES

### 3.1 The Decision Stages

Srinidhi & Vasarhelyi [19] discussed the usage of reliability theory for evaluating internal controls and identified the stages involved therein. Appendix I gives an overview of the technical aspects of the reliability method and presents the method of representing accounting activities and controls as components in reliability networks. The stages of decision making in internal control evaluation can be stated using reliability terminology as follows:

1. **Estimation of component reliabilities**: In this stage of decision making, the evidence collected on the accounting information system – such as the organization, the structure of activities and controls, the task complexity, etc. – are used in estimating the reliabilities of individual activities. These, in conjunction with the tests of compliance, are used to estimate control reliabilities.

2. **Aggregation of component reliabilities into system reliability**: The estimated component reliabilities are combined into system reliability numbers using the structure function.

3. **Interpretation of the system reliabilities in terms of the extent and timing of the substantive tests of detail.** Using prior judgments of materiality and tolerable audit risk level, the system reliability is mapped on to the degree of

---

[4] While the decision makers (auditors) need to be aware of the broad concepts on which these models are used, they need not be experts in the computational algorithms of these models to be able to use them. To give an example from a different area, investment analysts need not be experts in mathematics to be able to use the Black and Scholes [5] option pricing model in their decisions. They need to be aware only of the inputs to and the outputs from the model.

substantive testing.[5]

## 3.2 Estimation of Component Reliabilities

Reliabilities of activities and controls are influenced by factors such as organization, structure of activities and controls, personnel, task complexity, fatigue, overload, and performance evaluation methods. For instance, an organization with all its tasks segregated will have component tasks whose reliabilities can be expected to be higher than the components tasks of a system whose tasks are not segregated.[6]

In an organization with centralized hierarchy, the consequences of a detected error are likely to be severe and therefore, higher reliabilities can be expected. The influence of structure can be discerned if an activity without controls is compared with an activity with controls. Because of the preventive aspects of controls, the activity reliability can be expected to increase. The competence, the awareness and the integrity of the persons performing the tasks also influence component reliabilities. While it is not feasible to present strict mathematical relationship between the factors and component reliabilities, the nature of such relationship can still be studied. The expressions E1 to E3 {derived in [18]} relate the component reliability of a task to (i) certain global parameters such as hierarchy and the structure of activities and controls, (ii) task related parameters such as the benefit to the employee of an error in the task, (iii) auditor related parameters such as probability of detection and (iv) employee related parameters such as the subjective beliefs and utility of the employee. Table 1 gives the directions in which these parameters are influenced by the various factors.

This suggests that in estimating component reliabilities, the auditor could structure the estimation process after collecting evidence on factors by thinking in terms of the parameters identified above.

For a control procedure, component reliability is estimated as the product of the probability that it is applied and the probability that it is effective when applied. The evidence collected above and the parameters help in estimating the effectiveness of a control if it is applied.

---

5 The logical process is different. The system reliabilities can be "combined" to form prior beliefs on the audit risk level. Substantive tests provide the evidence needed to update these priors into posterior pdf of errors. In the auditing process, the same relations are useful in determining the degree of substantive testing needed to achieve the tolerable audit risk level.

6 A detailed study has been taken up to study the impact of duty segregation on components and system reliabilities [18]. In this paper, complete segregation of duties has been assumed.

Table 1: Factor Parameter Relationships

| Parameters-> Factors | $p_{uj}$ | $p_{1j}$ | $p_2$ | $b$ | $v_{1j}$ | $v_{2j}$ | $g$ |
|---|---|---|---|---|---|---|---|
| Segr.duties | -1 | | -1 | | | | |
| Hierarchy | 0 | 0 | 0 | | | | |
| Structure | 0 | | | 0 | | | |
| Competence | -1 | | | | | | |
| Awareness | -1 | | | | | | |
| Integrity | | | | | | | 0 |
| Task complexity | 1 | | | | | | |
| Fatigue | 1 | | | | | | |
| Rew.& Punish. | | | | 0 | 0 | 0 | |
| Perc.of audit | | 1 | 1 | | | | |

where: $P_{uj}$=Probability of an unintentional error

$P_{ij}$=Perceived probability of detection

$P_2$ =Cond. prob. that error is perceived intentional given that is detected

$b$ ="Benefit" to the employee of an undetected error

$v_{1j}$=Disut.to employee given that error is perceived intentional

$v_{2j}$=Disut.to empl.given that error is perceived unintentional

$g$ =Subjective Probability Distribution of the utility level at which employee will be be persuaded to intentionally commit an error.

Estimated Reliability of the task performed by an employee j

$$= p_j(p_u + [1-p_u]\, p_{ij}) \quad (E1)$$

where:

$$p_{ij} = \int_0^{u^1_j} g(x)dx\,, \quad (E2)$$

$$u^1_j = [(1-p_{1j})\int_b u_j f_j db] - p_{1j}[(p_2 v_{1j} + (1-p_2) v_{2j})] \quad (E3)$$

$u_j(b)$= utility function of the employee

$f_j(b)$= subjective probability distribution of benefit

Note: Cell values of 1 represent the potential of a positive relationship. Cell values of -1 represent the potential of a negative relationship. Cell values of 0 represent the potential of a positive or negative relationship. Blank signifies no relationship.

### 3.3 Aggregation of component reliabilities into system reliability

The second stage of decision making is the aggregation of the estimated component reliabilities into a unique system reliability number.

More specifically, let $\{X_1, X_2 ... X_n\}$ represent the states of the components with a structure function $\phi[X_1, ... X_n]$. Let the corresponding component reliabilities be $\{p_i\}$ such that

$E[X_i]=p_i$ for $i=1...n$. where E is the expectation operator

The system reliability is then given by

$$h[p] = E\{\phi[X_1 ... X_n]\}$$

This reduces to

$$h[p] = \phi[p_1 ... p_n].$$

if we assume that there are no dependencies.

Dependencies, however, can arise due to many reasons. Inter-transaction dependencies can exist if the underlying error-generation process is not a Bernoulli process. If material inter-transaction dependencies are expected by the auditor, his estimation can be made more accurate by dividing the audit period into sub-periods in which independence can be assumed. Statistical activity-control dependency can exist if the conditional reliability of the control given success of the activity is different from that given a failure of the activity. In these cases, it is possible to aggregate component reliabilities using "equivalent independent components". There also can exist magnitude based dependencies. It is common to have additional controls (or controls with higher reliability) for transactions involving higher values. This means that component reliabilities are to be treated as functions of transaction magnitude. In case of discrete cutoff points (in magnitude) at which the reliabilities change, different reliabilities can be computed for each magnitude range[7]. Identification of these dependencies and their incorporation into the decision framework is clearly a matter of subjective audit judgment. The contribution of the model in this stage is that it clearly separates the mathematical aggregation process from the audit decision process. Using the model, the auditor can now concentrate on the dependencies in the system and how they influence system reliability rather than performing

---

[7] A more detailed discussion is given in [18]

the pure aggregation function.[8]

### 3.4 Interpretation of the system reliabilities

In the interpretation stage, the auditor uses the system reliability numbers in conjunction with the tolerable audit risk and materiality to determine the extent and timing of substantive tests. In effect, he is interested in the way that system reliability numbers relate to the probability distribution of errors in account balances. The purpose of evaluating internal controls is to develop a prior expectation on the errors in the final balances. Appendix II. gives some "combination rules" which map on the error pdf's to system reliability numbers and then presents a "consolidation" method which uses the Bayesian updating process to integrate substantive test results with the prior distributions developed on the basis of internal control evaluation.

## 4 EMPIRICAL VALIDATION

### 4.1 Need for Empirical Study

The earlier sections identified the decision stages and the logical relationships in each one of the stages. A utility-based model was developed to structure the decision process of auditors in estimating component reliabilities. A reliability based model was presented for aggregating component reliabilities into a system reliability measure. A Bayesian approach was suggested to use the system reliability numbers in the audit process.

Of these three stages, the aggregation stage presents the opportunity for maximum improvement in audit judgment. A clearly established theory of aggregation of component reliabilities using a structure is available for use as a decision aid. It is not enough, however, to present the normative model. To demonstrate its usefulness in the audit situation, it is necessary to show that the audit decisions improve as a result of using this model. It is also necessary to show that the heuristics presently used by the auditors (in a descriptive sense) to aggregate evidence do not conform to the established relationships used in the model. If the normative relationships of the model also describe the current audit judgment process, the importance of the normative model would indeed be marginal.

This requires an empirical study to validate the model. An experimental methodology was used in this study. The characteristics used in this paper to study judgments are (i)

---

[8] Instead of "deciding" on system reliability using the component evidence, the auditor uses the model and gets the system reliability if there were no dependencies. He "anchors" his judgment at this point and modifies it based on the dependencies he identifies. One of the problems of decision processes is that decisions are often made on what are essentially known relationships. For example, one cannot "decide" on 3 + 4! So also, one cannot "decide" on the reliability of a system of independent components whose reliabilities are known.

the **consensus** in auditor judgment (ii) the **calibration** of auditor judgments with the evaluation based on the reliability model.[9]

## 4.2 Research Questions

The research questions addressed are the following:

1. Given the internal control structure and component reliabilities, to what extent does consensus exist among auditors in the judgment on substantive test restriction?

2. If the system reliability number is provided, does the consensus in the judgment improve and if so, by how much?

3. What is the degree of calibration of auditor judgments with the system reliability computed using the reliability model?

4. In evaluating the internal control system, what heuristics are used by the auditor in aggregating the evidence? Is adequate emphasis placed on the compensatory nature of activities and controls? (as suggested by the reliability model)

## 4.3 Research Methodology

Seventy-seven practising auditors took part in a laboratory experiment. The experimental task consisted of evaluating the Purchase Transaction Cycle of a hypothetical firm. The auditors taking part in the study were given a brief narrative description of the organization and a description of basic structural aspects of the Purchase transaction cycle. Standardized internal control documentation, was used to describe the experimental situation. Four major activities and controls of the cycle were presented at the two levels of reliability each. Auditors made judgments concerning: (1) the system reliability, and (2) substantive test restriction.

The purchase system presented to the auditor had three major activities - purchase ordering, receiving and vouchering - and two major controls. The representation of this system in a reliability framework is given in appendix IV.

Of the five procedures in the system, four were varied in the experiment. These four activities(controls) represent the factors in the experiment and the two levels of reliability, (at which each of these factors were presented) the factor levels. This is a $2^4$ factorial design. In the second section of the experiment the auditors gave rating of the

---

[9] Consensus refers to the agreement among auditor's responses to a given situation. Calibration refers to the closeness of auditor's response to the reliability computed using the model.

system when the system reliability numbers were provided to them. The ratings formed a graduated scale in terms of the extent and timing of substantive tests to be performed. Appendix III gives parts of the instrument and the rating scale used. Auditors give their judgments for each combination of such factor levels in terms of the system reliability and the degree of substantive testing. They constitute the dependent variables. The descriptive ANOVA technique is used for the discovery and description of factor usage.

### 4.4 Data Analysis

Measures: The measures of consensus between any two auditors (or between the auditor and the model) is the correlation coefficient between the responses of the two auditors. The overall measure of consensus is the coefficient of concordance.

The calibration of an auditor with the model is measured using both an associative and a distance measure.[10] The correlation between the reliabilities assessed by an auditor and the reliabilities computed by using the model provides an associative measure of the calibration of that auditor with the model. The difference between the model computed reliability and the auditor-elicited reliability (normalized by the total range of variation) is used as a measure of the distance. An appealing interpretation of these measures is to treat the assessed reliabilities as uncertain outcomes of a "perception" process based on the true normative reliability measures. In such a case, the bias measure is the intercept of a linear regression model of such a process. The association measure is the proportion of variance explained by such a model.

Data Analysis Design: Test-retest correlation is used to evaluate the data reliability [11] of system ratings and system reliabilities.

There is extensive literature on direct and indirect elicitation of subjective probabilities. [7]. In the experiment, both direct and indirect elicitation are used. Apart from giving error probabilities, the auditors also give degrees of substantive test restriction which can be "inverted" (using the corresponding function for each auditor) to yield indirect elicitation of system error probability judgments. The correlation between the directly and indirectly elicited reliabilities is computed for each auditor. A consistently good correlation across auditors can justify the use of directly elicited reliabilities and support the claim that splitting the decision into two stages does not produce adverse perceptual responses in auditors.

---

[10] consider two sets of numbers (.2, .05, .08) and (.4, .10, .16). There two sets are "perfectly associated". i.e. The relation between 2 and 4 is the same as the relation between 5 and 10, and 8 and 16. Consider the first set to be the model-computed reliabilities of 3 situations and the second set to be the auditor judgments. The correlation between the two is 1 but, the auditor's judgments are systematically different. This difference is captured by a distance measure.

[11] Not to be confused with component and system reliabilities

The calibration association and bias measures are computed for each auditor. Consensus is measured both by the coefficient of concordance (non-parametric measure) and by the mean inter-auditor correlation. The mapping of a system rating given the system reliability number represents a function and is designated as the "rating function" for each auditor. The rating function consistency is tested by an inter auditor correlation analysis of the ratings given by the auditors.

# 5 RESULTS

This section discusses the subjects used and relates the results of the experimental effort concerning: perception of system reliability, method of elicitation, auditor heuristics, the effect of subjective aggregation, calibration, sensitivity to component reliabilities and presents a summary of results.

## 5.1 Descriptive Statistics on Sample Characteristics

Table 2 gives the summary descriptive statistics on the sample. The sample is homogeneous in age, education and training, and consists of 77 professional auditors. Most of the auditors were about 25 years of age with 2 to 3 years experience[12] in auditing. All of them had undergone the basic training given in the firm up to the senior level. Almost all of them had accounting education either at the undergraduate or at the graduate level. A majority had experience in documenting and evaluating the internal control system in the purchase transaction cycle.

The subject selection procedure restricted the generalizability of the study's findings as they apply to a **single firm**. On the other hand this allows for a much more homogeneous subject sample and a common understanding of the specific meaning of internal control evaluation ratings. Most firms have their own internal control evaluation procedures, but these vary substantially among firms (Cushing & Loebbecke, 1983 [10]). Sample choice from different firms may introduce substantial ambiguity into the Ss' task or require thorough training of subjects in the specific internal control representation used in the instrument.

## 5.2 Perception of System Reliability

Table 3 gives the test-retest correlations. The overall test-retest correlations for both the ratings and the error probability judgments (0.759 and 0.649, respectively) are high and of comparable magnitude. Therefore, even a judgmental system reliability is perceived by the auditor to be as valid a measure of systems performance as the rating.

---

[12] Ashton (1974) [3] indicates that an auditor requires two and one half to three years of experience until being allowed to make this type of judgment. (p. 150)

Table 2: Summary Sample Statistics

| Variable | Descriptive Statistics | |
|---|---|---|
| Age | Range 23 to 37 | Median 25 |
| Sex | Males 52 | Females 23 |
| ICQ | Familiar 72 | Not familiar 5 |
| EXP | Mean 2.46yrs. | Median 2.5yrs |
| EXPTC | Experienced 57 | Not experienced 19 |

Variable Labels:

ICQ : Binary variable on familiarity with internal Control Questionnaire

EXPTC : Binary variable on experience in evaluating Purchase Transaction Cycle.

EXP : # of years of experience in auditing.

Table 3: Test Retest Correlations

| Sys.# | Correlations between ratings | | Correlations between reliabilities. | |
|---|---|---|---|---|
| | Pearson | Spearman | Pearson | Spearman |
| 1 | 0.75 | 0.737 | 0.465 | 0.670 |
| 2 | 0.863 | 0.910 | 0.750 | 0.845 |
| 3 | 0.851 | 0.836 | 0.598 | 0.542 |
| 4 | 0.873 | 0.879 | 0.882 | 0.880 |
| 5 | 0.555 | 0.328 | 0.256 | 0.786 |
| 6 | 0.752 | 0.741 | 0.443 | 0.435 |
| 7 | 0.636 | 0.641 | 0.721 | 0.652 |
| 8 | 0.589 | 0.436 | 0.803 | 0.781 |
| 9 | 0.371 | 0.305 | 0.240 | 0.404 |
| 10 | 0.524 | 0.531 | 0.762 | 0.623 |
| 11 | 0.592 | 0.594 | 0.354 | 0.279 |
| 12 | 0.500 | 0.471 | 0.671 | 0.709 |
| 13 | 0.311 | 0.375 | 0.479 | 0.556 |
| 14 | 0.288 | 0.271 | 0.613 | 0.600 |
| 15 | 0.510 | 0.471 | 0.534 | 0.393 |
| 16 | 0.721 | 0.755 | 0.628 | 0.673 |
| Overall | 0.759 | 0.751 | 0.649 | 0.715 |

## 5.3 Method of Elicitation

Table 4 gives the association between directly and indirectly elicited system reliabilities. A mean correlation of 0.834 and a 95 percentile range of 0.71 - 0.986 indicates a consistently good correlation across the auditors. The results will not materially differ if rank correlations are used in the analysis.

The numbers also confirm the feasibility of using probability measures in internal control evaluation. There is no substantial difference between the product moment correlations and rank correlations. This indicates that the result is not driven by a few extreme cases.

**Table 4:** Direct - Indirect Reliability Judgment Correlations

| | PEARSON | SPEARMAN |
|---|---|---|
| MEAN | 0.8336 | 0.8557 |
| STANDARD DEVIATION | 0.1776 | 0.1696 |
| STD. ERROR OF MEAN | 0.020 | 0.019 |
| 95 PERCENTILE RANGE | 0.71-0.986 | 0.722-1.00 |

## 5.4 Auditor Heuristic

Table 5 gives the results of ANOVA with assessed system reliabilities as dependent variables and component reliabilities as explanatory variables. The main effects are predominant over the interaction effects indicating that auditors use a linear heuristic to aggregate component reliabilities.

This means that auditors do not give adequate weightage to the compensatory nature of activities and controls. The reliability model suggests a much higher interaction effect. This, in combination with the low calibration displayed by auditors, supports the case for using the reliability model to improve their judgments.

## 5.5 The Effect of Subjective Aggregation

Table 6 presents the consensus among auditors both when only component reliabilities are given and when system reliability is also given. Clearly, the consensus is very low when only component reliabilities are known. The consensus improves dramatically when system reliability numbers are provided. The coefficient of concordance, which was 0.5566 when only component reliabilities are given, improves to 0.9465 when system reliability is given. This improvement is highly statistically significant [t = 146.3]. The magnitude of the improvement in consensus achieved here is a measure of the variance in judgments caused in aggregation process.

Table 5: Analysis Of Variance On System Reliabilities

Dependent Variable: System Reliability

| Factors | Levels | |
|---|---|---|
| V1: Preparation & Review of P.O. | .7 | .98 |
| V2: Comparison of VI to P.O. & R.R. and approval of VI. | .65 | .98 |
| V3: Voucher Preparation | .8 | .99 |
| V4: Comparison of Voucher to VI, P.O. and R.R. | .75 | .98 |

# of obs: 77 auditors*16 systems = 1232

| | S.S | DF | M.S. | F |
|---|---|---|---|---|
| MAIN EFFECTS | 2.1641 | 4 | .541 | 90.17 |
| INTERACTION EFFECTS | .0542 | 11 | .0049 | .821 |

INDIVIDUAL MAIN EFFECTS

| VARIABLE | S.S = M.S | F |
|---|---|---|
| V1 | .1845 | 30.15 |
| V2 | 1.0796 | 176.38 |
| V3 | .2480 | 40.51 |
| V4 | .6516 | 106.45 |

### 5.6 Calibration

Table 7 gives the calibration association and calibration bias measures. The mean calibration association level is 0.549, with a range from 0.233 to 0.878, with a standard deviation of 0.154. The normalized calibration bias measure has a mean of 0.4079, which indicates significant underestimation. In fact, 25 of the 77 auditors consistently underestimated all reliabilities and no auditor consistently overestimated all reliabilities. These results show that auditors are poorly (but positively) calibrated for the model and significantly underestimate system reliability.

### 5.7 Sensitivity to Component Reliabilities

An attempt was made to study the sensitivity of the system reliability to the component reliabilities. Table 8 gives the results of the sensitivity analysis. Auditors are more sensitive to decreases in component reliabilities than the model prescribes. Further probe into the auditor judgment pattern was made using principal component analysis. The principal component analysis revealed that auditors gave greater importance to the vouchering stage than the ordering and receiving stages in the purchase cycle. There was no clear clustering of the auditors in the two-factor space.

Table 6: Consensus Data

# OF AUDITORS = 77

| | RATINGS | RELIABS. | RATINGS given sys. reliabs. |
|---|---|---|---|
| 1. COEFFICIENT OF CONCORDANCE | 0.5566 | 0.5264 | 0.9465 |
| 2. MEAN PEARSON COEFFICIENT | 0.5758 | 0.4835 | 0.9394 |
| 3. ESTIMATED* STD. DEV. | 0.0812 | 0.1901 | 0.044 |
| 4. STD. ERROR OF THE MEAN | 0.0093 | 0.0135 | 0.0032 |
| 5. RANGE OF AVG. CORR. COEF. | 0.260-0.718 | 0.0316-0.631 | 0.8265-0.9634 |
| 6. MEAN SPEARMAN COEFFICIENT | 0.5508 | 0.5202 | 0.9458 |
| 7. ESTIMATED * STD. DEV. | 0.0864 | 0.1118 | 0.0346 |
| 8. STD. ERROR OF THE MEAN | 0.0098 | 0.012 | 0.0027 |
| 9. RANGE OF THE CORR. COEFF. | 0.274-0.691 | 0.017-0.658 | 0.8394-0.9623 |

* Mean of std.deviations of correlations for each auditor

Table 7: Calibration Association Measures

| | ASSOCIATION | | | | BIAS | |
|---|---|---|---|---|---|---|
| | MODEL-DIRECT | | MODEL-INDIRECT | | NORMAL-IZED | NORMAL-IZED |
| | PEARSON | SPEARMAN | PEARSON | SPEARMAN | | |
| MEAN | 0.5408 | 0.6476 | 0.5487 | 0.6615 | 0.0616 | 0.4079 |
| S.DEV. | .1694 | .1567 | .1541 | .1453 | 0.056 | 0.3708 |
| SE OF MEAN | .019 | .018 | 0.018 | .017 | 0.006 | 0.042 |
| T VALUE | | | | | 9.65 | 9.65 |
| SIG.PROB. | | | | | <.01 | <.01 |
| RANGE | -.032 TO .83 | -.012 TO .89 | .233 TO .878 | .272 TO .917 | | |

Table 8: Sensitivities Of Components

| Procedure | Col1 DSR | Col2 DCR | Col3 SEN | Col4 MES | Col5 SEN/MES |
|---|---|---|---|---|---|
| 1. Prep. & Review of Purchase Order | .1867 | .28 | .67 | .17 | 3.94 |
| 2. Comparison of VI with PO, RR & approval | .4485 | .33 | 1.36 | .16 | 8.5 |
| 3. Voucher Preparation | .2455 | .19 | 1.29 | .13 | 9.92 |
| 4. Comparison of Voucher with VI, PO & RR | .3735 | .23 | 1.62 | .10 | 16.2 |

LEGEND:

DSR: Decrease in System Reliability.
DCR: Decrease in Component Reliability.
SEN: Sensitivity.
MES: Sensitivity as computed by the model.

## 5.8 Summary of Results

The following most significant results were obtained from the experiment:

1. Auditors perceive system reliability to be as valid a measure of evaluation as the currently used ratings methods. The directly and indirectly elicited system reliability judgments are highly associated. This confirms the feasibility of reliability measure in system evaluation.

2. The descriptive ANOVA results indicate that auditors placed a lower emphasis on interactions between activities and controls than the model suggested. The compensatory nature of activities and controls was not recognized. This supports the case for using reliability model as a decision aid.

3. The auditors displayed a low level of consensus in aggregating component reliabilities into a system reliability measure. They also displayed very poor calibration with the model. There was a significant underestimation of system reliability. These findings lead to the conclusion that auditors do not display expertise in aggregating component reliabilities into a system reliability measure.

The above results indicate that using the reliability model in internal control evaluation is feasible and may be useful in improving the quality of auditor decisions.

# 6 CONCLUSION

The purposes of this study were twofold: (1) to develop a probability based model for aiding auditors in evaluating Internal control systems and (2) to present empirical results on how the use of such a decision aid might influence both the decision process and the actual decisions of auditors.

The motivation for developing a quantitative model was that by breaking up a complex decision process into simple decision stages, by systematizing the nature of judgment at each stage, by aiding this judgment with known quantitative relationships and by documenting judgments in numbers, an "improvement" in decisions could be brought about. Earlier models and the convenience of the probability measure led to the use of a probability model.

Concepts from Reliability theory were used in developing the evaluation model. The accounting system was viewed as a network of activities and controls through which transactions were processed to yield financial statements. Activities and controls were represented as components characterized by a reliability measure. The interplay of activities and controls in handling transactions was captured by the "Structure Function". This model differed from the earlier models in representing the components by single reliability numbers.

The reliability network model was then applied to the decision process of auditors. The constraints of viewing accounting systems as reliability networks on audit judgment were discussed. The determinants of reliabilities and the dependencies were sought to be incorporated into the model. Auditing was viewed as a sequential decision making process and the stage in which this conceptualization of accounting systems would aid auditors was identified. Possible impact and use of the outcomes of reliability model aided evaluation on other audit decision stages was presented.

The main contribution of the analytical part of this study is in identifying the possible influence on the audit decision process of conceptualizing accounting systems in reliability terms.

Empirical evidence was then collected on the influence of the use of reliability decision aid on auditor judgments in terms of consensus and calibration. The presence or absence of consensus and calibration under different conditions of auditor decision making was related to the possible use and impact of the reliability model. More specifically, auditors were found to be *poorly calibrated* and *lacked consensus* in their system judgments when the component reliabilities were provided. The empirical evidence supports the use of reliability model in audit evaluation of internal control systems.

# I. THE RELIABILITY MODEL

## I.1 TECHNICAL ASPECTS - AN OVERVIEW

The reliability of a component is the probability of its successful operation [under prescribed conditions] as a part of the system. Similarly, the reliability of a system is the probability with which the system performs successfully. Implied in these definition is a clear objective meaning of "success".

For both systems and components only two states are distinguished. To indicate the state of a component, i, a binary variable $x_i$ is assigned such that

$x_i = 1$ if the component i is successful
$= 0$ otherwise.

The state of the system is represented by a binary variable $\phi$ such that
$\phi = 1$ if the system is successful
$= 0$ otherwise.

By definition, the state of a system is the function of the states of its components.
i.e., $\phi = \phi(X_1, X_2, \ldots, X_n) = \phi(X)$
where $X = (X_1, X_2, \ldots, X_n)$.

$\phi(X)$ is defined as the STRUCTURE FUNCTION of the system. N, the number of components, is called the ORDER of the system.

In these terms, the component reliability $p_i$ is the expected value of the random variable $N_i$ and the system reliability function $h[P] = E\{B(X)\}$ where $P = (P_1, P_2, \ldots, P_n)$ and E is the expectation operator.

The structure of a system which cannot perform successfully if even one of the components fails is defined to be a SERIES STRUCTURE. The structure function of such a system is given by

$$\phi(X) = N_1 * N_2 * N_3 \ldots X_n.$$

It is easily seen that if $x_i = 0$ for any i,

$$\phi(X) = 0.$$

If the components are independent,

$$h(P) = E[\phi(X)] = E(X_1; X_2, \ldots, X_n) = E(X_{1*}E(X_{2*} \ldots E(X_n))$$
$$= P_1 P_2 \ldots P_n.$$

Therefore, the system reliability of independent components in series is equal to the product of their reliabilities.

The structure of a system which performs successfully if even one of its components is successful is defined to be a PARALLEL STRUCTURE. The structure function of such a system is given by: $\phi(X) = 1-(1-X_1)(1-X_2)(1-X_3)...(1-X_n)$. It is easily seen that if $N_a = 1$ for any i, $\phi(X) = 1$.

If the components are independent,

$$h(P) = E[\phi(X)] = E[1-(1-X_1)...(1-X_n)] = 1-\{1-E(X_1)\}\{1-E(X_2)\}...\{1-E(X_n)\}$$

$$= 1-(1-p_1)(1-p_2)...(1-p_n).$$

Every system can be represented by a combination of series and parallel structures. Therefore, the reliability of any system of independent components can be represented as a function of the reliabilities of the components. Another interesting result from reliability engineering is that system reliability is bounded from above and below and these bounds are functions of component reliabilities irrespective of whether they are independent or not. For a more complete discussion, refer [18].

## 1.2 REPRESENTING AN ACCOUNTING SYSTEM AS A STRUCTURE OF COMPONENTS

In an accounting system, source documents are produced to record events of financial relevance. The first critical recording of economic signals from an event is referred to as an 'ACTIVITY'. A procedure by which errors in activity performance can be reduced is called a CONTROL.

The success of an activity is defined as the valid identification of the event and its error-free documentation.

The success of a control is defined as a successful application of that control. The reliability of a control is the joint probability that the control is applied (compliance) and that it is effective when applied. For an error to occur in the output of a sub-system with one activity and controls on that activity, there should be an error in the activity as also all the controls of that activity. This is analogous to the logic of a parallel structure. Therefore, controls are represented by components placed in parallel with the activity.

A control may reduce errors in more than one activities. In such a case, the control component is placed in parallel to all the activities controlled for. In this manner, all the activities and controls in an accounting system can be represented by a network of interrelated components.

Adapting such a logic to accounting systems implies that (i) the state of the system is determined completely by the individual activities and controls in the system and that (ii) the system reliability is a meaningful evaluative measure. The first implication seems to constrain the evaluation to only those controls which can be represented as reliability components. Controls such as segregation duties and variables such as personnel competence are treated as determinants of component reliabilities and are thus incorporated in the evaluation. In evaluating internal control systems, the auditor is interested primarily in error frequencies. (SAS 39) Reliability is a measure of error frequencies and is therefore a meaningful measure of internal control strength. Therefore, the internal control aspects of an accounting system satisfy the premises required for using combinatorial logic.

It may be noted here that the time-sequence of operations which forms the essential feature of a flowchart is not of prime importance in reliability network representation.

Representation of activities and controls in a reliability network also requires an optimum degree of aggregation of procedures. For example, the activity "preparation of purchase order" is, in fact, an aggregation of many micro-activities such as typing, mailing, correcting typing errors, etc. The degree of detail need not be more than the actual detail of documentation in the system.

# II. COMBINATION AND CONSOLIDATION

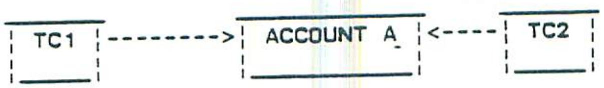
## II.1 COMBINATION RULES

A "combination rule" as presented here is a mathematical expression which relates the parameters of the probability density function of errors in a ledger account and the internal control reliabilities of the transaction cycles which influence those ledger accounts.

### II.1.1 Combination of transaction cycle reliabilities to yield the probability of there being no material error in an account balance.

Consider an account A which is influenced by two transaction cycles TC1 and TC2. [See fig.II-1]

Figure II-1: Transaction Cycles and Account




Let the estimated number of transactions over the period of the audit for TC1 be $N_1$ and for TC2 be $N_2$. Let the reliability of TC1 be $P_1$ and of TC2 be $P_2$. Let the errors be distributed as $f(e1)=N(u_{e1}, s_{e1}^2)$ in TC1 and as $f(e2)=N(u_{e2}, s_{e2}^2)$ in TC2. where N(.,.) represents a normal distribution.

Joint probability density of there being an error of magnitude e in TC1 $=(1-P_1)f(e1) = Q_1 f(e1)$ where $Q_1 = (1-P_1)$.

Expected mean total error in account A due to TC1 is

$$\sum_{n=0}^{N_1} \binom{N_1}{n} Q_1^{\,n} P_1^{\,N_1-n} (nu_{e1})$$

$$=N_1 Q_1 u_{e1}$$

$$\sum_{n=0}^{N_1} \binom{N_1}{n} Q_1^{\,n} P_1^{\,N_1-n} nu_{ei}$$

Expected variance of the error in Account A due to TC1 is

$$\sum_{n=0}^{N_1} \binom{N_1}{n} Q_1^n P_1^{N_1-n} (n s_{e1}^2)$$

$$= N_1 Q_1 s_{e1}^2$$

The error due to TC2 will have a mean $N_2 Q_2 u_{e2}$ and a variance $N_2 Q_2 s_{e2}^2$.

The error in financial statement account A will be normally distributed with mean $N_1 Q_1 u_{e1} + N_2 Q_2 u_{e2}$ and variance $N_1 Q_1 s_{e1}^2 + N_2 Q_2 s_{e2}^2$.

In general, if the account is influenced by m transaction cycles and the normality assumption holds, the total error will be distributed as $N(U_E, S_E^2)$

where $U_E = \sum_i N_i (1-P_i) u_{ei}$

and $S_E^2 = \sum_i N_i (1-P_i) s_{ei}^2$

If the materiality limits are $(-M_1, +M_2)$ for the error, the probability of there being no material error in A is given by:

$$\int_{-M_1}^{M_2} N(U_E, S_E^2)\, de$$

### II.1.2 Mean and variance of the pdf of errors when system reliability estimates are beta distributed and errors are normally distributed.

Consider 2 transaction cycles as before influencing an account A.

Reliability of TC1 ~ $f_1(P_1)$ and

Reliability of TC2 ~ $f_2(P_2)$ where $f_1$ & $f_2$ are beta distributions.

Expected value of total error in Acc. A due to TC1 is

$$\int_0^1 \sum_{n=0}^{N_1} \binom{N_1}{n} Q_1^{\,n} P_1^{\,N_1-n} f_1(P_1)(nU_{e1})dP_1$$

$$= \int_0^1 n\binom{N_1}{n} Q_1^{\,n} P_1^{\,N_1-n}\, U_{e1} f_1(P_1)dP_1$$

$$= \int_0^1 N_1 Q_1 U_{e1} f_1(P_1)dP_1$$

$$= N_1 U_{e1} \int_0^1 Q_1 f_1(P_1)dP_1$$

$$= N_1 U_{e1} [1 - a/(a+b)] = N_1 U_{e1} [b/(a+b)]$$

where a and b are the parameters of the beta distribution.

The Variance $V(e_1)$ is given by:

$$\int_0^1 \sum_{n=0}^{N_1} \binom{N_1}{n} Q_1^{\,n} P_1^{\,N_1-n} f_1(P_1)(nS_{e1}^{\,2})dP_1$$

$$= \int_0^1 [N_1 P_1(1-P_1) + N_1^{\,2}(1-P_1)^2 \; S_{e1}^{\,2}]\; f_1(P_1)\, dP_1$$

This expression reduces to the following:

$$N_1 S_{e1}^{\,2}[N_1 + (1-2N_1)a/(a+b) + (N_1-1)\{V(P_1)+a^2/(a^2+b^2)\}]$$

where V(.) represents the variance operator.

As before, if the account is
influenced by m transaction cycles, the total error
will have

$$\text{Mean} = \sum_{i=1}^{m} N_i u_{ei} b_i / (a_i + b_i) = U_E$$

and

$$\text{Variance} = \sum_{i=1}^{m} V(e_i)$$

where $a_i$ and $b_i$ are the parameters of the beta distribution corresponding to the reliability of the transaction cycle i and $V(e_i)$ is the variance of the error because of the transaction cycle i.

## II.2 CONSOLIDATION

The consolidation method presents a Bayesian updating algorithm treating the outcome from a combination rule as the prior distribution of errors in a ledger account balance and using direct evidence from substantive tests.

### II.2.1 Case 1

If the prior probability distribution is a discrete one characterized by

$\text{Prob}(E_i) = P(E_i) \quad i = 1,2, - - - m$, and

$$\sum_{i=1}^{m} P(E_i) = 1,$$

and the account is divided into n sub-accounts of which k are tested. Let the error in the tested sub-account be $e_i$ where i = 1, 2...k.

Posterior prob. of $E_i$ after considering e,

$$= \left[ \frac{P(e_1/E_i)P(E_i)}{\sum_{j=1}^{m} P(e_1/E_j)P(E_j)} \right]$$

If the updating operation can be represented by u, then,

$$P_{post}(E_i) = u^k\{ P(E_1), P(E_2), \ . \ . \ . \ P(E_m) \}$$

### II.2.2 Case 2

If the probability distribution is a continuous distribution f(E) with no "probability mass" at any point,

$$\text{Posterior dist.} = \frac{P(e_1/E)*f(E)}{\int_E P(e_1/E)*f(E)dE} = f_{post1}(E)$$

$$= u\{f(E)\}$$

$$\text{Final posterior distribution} = u^k\{f(E)\}$$

### II.2.3 Case 3

If the prior probability distribution has a probability mass $P_0$ at 0 and is a continuous distribution f(E) at all other points,[13]

$$\text{Post. prob. mass at } 0 = \frac{P(e_1/0)*P_0}{P(e_1/0)*P_0 + \int_E P(e_1/E)*f(E)dE}$$

$$=u_1(f,P_0)$$

$$f_{post1}(E) = \frac{P(e_1/E)*f(E)}{P(e_1/0)*P_0 + \int_E P(e_1/E)*f(E)dE}$$

$$=u_2(f,\ P_0)$$

Final posterior mass at 0 is $u_1^k(f,P_0)$ and

the final posterior distribution is $u_2^k(f,P_0)$ .

[13] This is reasonable because we have a number of transactions which are error-free giving rise to the "mass"

# III. INSTRUMENT SUMMARY

1. A brief narrative of the background, organization, accounting system, and design of purchase transaction cycle using ICQ worksheets.

2. Questionnaire.

3. Section 1

   Given: Accuracy of performance of each accuracy and control.

   Asked: System rating - point estimate
   System rating - range estimate
   System reliability - circle one of the given numbers.

   Section 2

   Given: System reliability

   Asked: System rating.

4. Rating Scale:

   0 No reliance and timing restricted to the year end.
   1 Low reliance but timing restricted to the year end.
   2 Low reliance with timing restricted to within 1 month of year end.
   3 Moderate reliance with timing restricted to within 1 month of year end.
   4 Moderate reliance with timing restricted to within 2 months of year end.
   5 High reliance with timing restricted to within 2 months of year end.
   6 High reliance with timing allowed to be more than 2 months from year end.

# IV. ILLUSTRATION OF THE RELIABILITY FRAMEWORK

Figure IV-1 presents the reliability network of the purchase system used in the experiment. This figure was not presented to the participants.

**Figure IV-1:** Reliability Network of the Purchase System

```
    X1                        X2                   X4
|---o-------------------------o-------|    |-----o------|
|Preparation and         Preparation of|    |  Entry in  |
|Review of Purchase      Receiving     |    |  Voucher   |
|Order (PO)              Report (RR)   |    |  Register  |
|                                      |----|            |----
|                                      |    |            |
|                  X3                  |    |    X5      |
|-----------------o--------------------|    |-----o------|
 Comparison of Vendor's Invoice (VI) with   Comparison of
 the PO and RR and approval of VI           the voucher with
                                            the support
                                            documents
```

Structure function $\phi = [1-(1-X_1X_2)(1-X_3)] * [1-(1-X_4)(1-X_5)]$

Note on X1: X1 can be presented as an activity with a control in parallel. The advantage of a reliability method is to enable the evaluators to choose the level of detail they want to represent. X1 can also be presented as many activities - preparation of draft p.o., approval, typing, mailing out the copies etc... or as one procedure as shown here.

# REFERENCES

1. AICPA. Codification Of Statements On Auditing Standards. AICPA, New York., 1980.

2. Ashton, Robert. *Human Information Processing.* American Accounting Association, 1982.

3. Ashton, Robert. "An Experimental Study Of Internal Control Judgments". *Journal Of Accounting Research 59* (1974), 143-157.

4. The Committee on Auditing Concepts. *A Statement of Basic Auditing Concepts.* American Accounting Association, 1973.

5. Black, F. and Scholes, M.S. "The Pricing of Options and Corporate Liabilities". *Journal of Political Economy* (June 1973), 637-654.

6. Bodnar, G. "Reliability Modeling of Internal Control Systems". *Accounting Review 50* (1975), 747-757.

7. Chesley, G.R. "The Elicitation of subjective probabilities.". *Journal Of Accounting Research 61* (1976), 27-48.

8. Cushing, Barry E. "A Further Note On Mathematical Approach To Internal Control". *Accounting Review 50* (1975), 151-165.

9. Cushing, Barry E. "A Mathematical Approach To The Analysis And Design Of Internal Control Systems". *Accounting Review 49* (1974), 24-41.

10. Cushing, Barry E. and Loebbecke, James K. Comparison Of Audit Methodologies Of Large Accounting Firms. University Of Utah, July, 1983. Unpublished Working Paper.

11. Cushing, Barry E. and Loebbecke, James K. "Analytical Approaches to Audit Risk: A Survey And Analysis". *Auditing: A Journal of Practice & Theory 3*, 1 (Fall 1983), 23-41.

12. Einhorn. "Expert Judgment: Some necessary conditions and an example.". *Journal of Applied Psychology 59* (October 1974), 562-571.

13. Goldberg, L.R. "Simple Models or Simple Processes? Some Research on clinical judgment.". *American Psychologist* (July 1968), 491.

14. Grimlund, Richard A. "An Integration Of Internal Control System and Account Balance Evidence". *Journal Of Accounting Research 20* (Autumn 1982), 316-342.

15. Hamlen, S.S. "A chance-constrained mixed Integer-programming Model for Internal Control Design.". *Accounting Review 55* (October 1980), 578-590.

16. Ishikawa, A. "A Mathematical Approach To Analysis And Design Of Internal Control : A Brief Comment.". *Accounting Review 50* (January 1975), 148-150.

17. Joyce, Edward. "Expert Judgment in Audit Planning". *Supplement to the Journal Of Accounting Research 14* (1976), 29-60.

18. Srinidhi, B. N. *Probability Modeling Of Internal control systems In Audit Decision Making*. Ph.D. Th., Columbia University Graduate School Of Business, 1983.

19. Srinidhi,B.N. & Vasarhelyi,M.A. Auditor Judgment Concerning Establishment of Substantive Tests Based on Internal Control Reliability. Columbia University, 1985. Unpublished Working Paper.

20. Srivastava. Reliability Modeling Of Internal Control Systems. PhD dissertation in progress, University Of Oklahama.

21. Stratton, William O. *Acccounting Internal Control Systems: Their Reliability and dichotomic structure functions*. Ph.D. Th., Claremont Graduate School, 1977.

22. Stratton, William O. "Accounting Systems: The Reliability Approach To Internal Control Evaluation". *Decision Sciences 12* (1981), 51-67.

23. Yu. S. and Neter, J. "A Stochastic Model Of The Internal Control System". *Journal Of Accounting Research 2* (1973), 273-295.