1-1-1986

Unresolved issues in classical audit sample evaluations

Donald R. Nichols

Rajendra P. Srivastava

Bart H. Ward

Recommended Citation

Auditing Symposium VIII: Proceedings of the 1986 Touche Ross/University of Kansas Symposium on Auditing Problems, pp. 105-119;

4

Unresolved Issues in Classical Audit Sample Evaluations

Donald R. Nichols
Texas Christian University

Rajendra P. Srivastava
The University of Kansas

Bart H. Ward
The University of Oklahoma

Classical variables techniques can be usefully employed in certain audit situations. They may be useful, for example, when auditing high error rate populations or accounts with numerous negative balances or when the auditor is concerned about both over and under-statement errors. Classical variables techniques may also be useful when the auditor is concerned with assessing the reasonableness of proposed adjustments in light of statistical test results. This paper reviews several issues associated with the evaluation of classical statistical hypothesis testing results in auditing. Though presented in terms relevant to classical statistical testing, some of the issues reviewed may be germane to other statistical or non-statistical approaches to audit sampling as well.

Some of these issues have been isolated and examined in greater detail by other studies. This paper mainly deals with the comparison and reconciliation of certain alternative evaluation strategies which can be employed when achieved allowances for sampling risk differ from planned levels. This situation can occur when the apparent achieved efficiency of a sample estimator is different from the level on which the auditor based the audit sampling plan.

Comparative Evaluation Strategies

Several strategies are available for use in evaluating the results of a classical variables hypothesis test. Conclusions drawn from the evaluation of sample results may vary depending upon which strategy is employed. Three of these strategies are explained and compared in this paper.

No one of the three strategies is uniformly dominant or necessarily superior to the others in all situations. However, they can lead to different conclusions. Therefore, it is important to understand how they differ. In this respect, the selection of an appropriate evaluation strategy is similar to the dilemma encountered in selecting an appropriate error bound in probability-proportional-to-size sampling applications (see Felix, Leslie & Neter, 1982).

We shall identify the three strategies as: the acceptance risk control strategy, the rejection risk control strategy, and the balancing strategy. Figure 1 depicts and compares the decision sequences associated with the first two strategies. The decision sequence for the balancing strategy is presented separately in Figure 2. Where possible the symbols and terminology used will conform to the AICPA audit sampling literature (e.g. Roberts, 1978; SAS 39; and Accounting and Audit Guide—*Audit Sampling,* 1983).

Both approaches described in Figure 1 are relatively well-known strategies for evaluating the results of classical statistical samples. Evaluation strategies based on both approaches appear in the AICPA's publication *Audit Sampling* as well as auditing literature and firm procedure manuals.

The acceptance risk control strategy for evaluation of the results of classical variable hypothesis tests appears in the AICPA publication, *Statistical Auditing* [Roberts, 1978]. This approach is also referred to, but not described in detail, by the AICPA's guide on audit sampling which supports Statement on Auditing Standards #39 (SAS 39) [Auditing Standards Board, 1981]. Sample evaluation approaches based on this strategy can be found in the auditing literature, e.g. Guy and Carmichael [1986].

The rejection risk control strategy is described in detail in the AICPA publication *Audit Sampling* and the audit and accounting guide prepared by the Statistical Sampling Subcommittee to support SAS 39. Sample evaluation approaches based on this approach can be found in the accounting literature, e.g. Arens and Loebbecke [1981] and Bailey [1981].[1]

The balancing strategy which is depicted in Figure 2 was explored by Thompson [1982] and is rooted in the work on the utility of various schemes for reporting or summarizing hypothesis testing results done by Leamer [1978]. Using this balancing strategy, the auditor would employ an epistemic loss minimizing criterion. It could be used as an alternative to the two better known traditional strategies.

In order to set the stage for the sample evaluation strategies portrayed in Figures 1 and 2, it may be useful to briefly consider the sample planning process. In most descriptions of audit sampling, in the planning stage, sample sizes are determined which will control the risk of incorrect acceptance (TD) and the risk of incorrect rejection (α) to levels that are acceptable to the auditor given *ex ante* (before sampling) information about the population and planned statistical estimator. In this regard, the estimated standard error is important.

The *ex ante* (planned) allowance for sampling risk associated with the amount A can be compiled based on an estimate of the standard deviation of the population under examination or the related population of auxiliary values (differences or ratios between audited and book values, etc.) and on auditor decisions about appropriate levels for the risks of incorrect acceptance and incorrect rejection and about the amount of tolerable error for the account or balance, TE. Discussions of this process and factors affecting it can be found in the audit sampling literature, especially Guy and Carmichael [1986], Arens and Loebbecke [1981], Roberts [1978], SAS 39 and the associated AICPA audit guide. The auditor will plan a sample such that if $B \in \hat{X} \pm A$, the reported amount will be accepted as fairly presented, whereas if B is not in the interval $\hat{X}$

$\pm$ A, the reported amount will not be accepted as fairly presented. In each instance, B is the book value of the account or balance and $\hat{X}$ is the audit sampling estimator of the correct value. The sampling plan will be established such that the risk of incorrect acceptance and the risk of incorrect rejection of the decision interval, $\hat{X} \pm A$, in relation to TE will be at levels planned by and acceptable to the auditor.

As shown in Figures 1 and 2, if the *ex post* (after sampling) information agrees with the *ex ante* estimates ($A' = A$), then the auditor faces no special evaluation problem, and the three evaluation schemes are the same. That is, the decision rule is to accept the book value if $B \in \hat{X} \pm A'$. Since this is equivalent to $B \in \hat{X} \pm A$, the associated risks of incorrect acceptance and rejection should be the same as the planned levels. A′ is the monetary amount which equates the risk of incorrect rejection associated with this decision rule with the planned level for α.

Usually, however, after the sample has been selected and audited, the *ex post* assessment of the standard error will be different from the *ex ante* assessment, i.e., $A' \neq A$. When this is the case, no decision strategy discussed in the auditing literature reviewed here will retain the risk of incorrect acceptance and the risk of incorrect rejection at the planned levels. For any given sample result, there is a trade-off between the two risks. In fact, there are infinitely many α and TD risk level pairs that could be established for the sample evaluation. In this circumstance, the issue to resolve is how to devise an evaluation strategy which will contain risk levels which are preferable or acceptable to the auditor. The three strategies discussed here handle the balancing of these risks in different ways. By understanding the approach and the results of these strategies, the auditor may select one (or devise another strategy) that is consistent with his or her preferences.

The essential differences among all the strategies reviewed in Figures 1 and 2 can be traced to different philosophies about risk control. In our discussion we shall highlight the manner in which each strategy deals with this dilemma and attempt to explain what the various options imply about the relative utility of incorrect rejection and incorrect acceptance.

Acceptance Risk Control

The acceptance risk control strategy, as detailed by Roberts [1978], will be reviewed first. Like each of the other strategic options discussed, the principal purpose is to provide a framework for rational evaluation of a classical statistical sample. The objective is to accept or reject the amount being tested, given the *ex ante* specification of the risk of incorrect acceptance, TD, and risk of incorrect rejection, α, and the achieved sampling test results.[2]

If the estimated standard deviation used in planning and the sampling estimator of standard deviation are identical, then the potential for variability in sampling results can be properly controlled by relying on the critical limits associated with the *ex ante* allowance for sampling risk. In such instances, $A' = A$, and an appropriate decision rule is to accept the amount being tested B, if $B \in \hat{X} \pm A'$. Otherwise it is appropriate to reject the amount being tested. In this situation, the planned risks of incorrect rejection and incorrect acceptance are also the levels achieved.

In most instances *ex post* estimates of the standard error of the estimator will vary from planned levels, i.e., $A' \neq A$. In pursuing the "acceptance risk control" option as shown in Figure 1, the auditor confronted with a difference between *ex ante* and *ex post* estimates of variability will establish an *ex post* allowance for sampling risk by relying on an initial decision rule which calls for acceptance of the amount under examination if that amount exists in the region $\hat{X} \pm A''$. In this case, A″ is the monetary amount which necessarily equates the risk of incorrect acceptance associated with the new decision rule (TD′) with the planned level for TD. In other words, the risk of incorrect acceptance associated with the decision interval $\hat{X} \pm A''$ is equivalent to the level originally planned by the auditor. The acceptance risk control approach does not explicitly control the risk of incorrect rejection. At this point, the risk of incorrect rejection may be higher or lower than the planned level, α. In other words, TD is fixed at the planned level and α varies, either higher or lower than the planned level.

The strategy as described so far can only lead to an acceptance decision where $B \in \hat{X} \pm A''$. The preeminence of TD is justified because at this point the initial decision rule allows only for acceptance of the reported amount. If acceptance is not possible, then rejection based on statistical evaluation alone cannot take place without considering the level of control over the risk of incorrect rejection.

In fact, if the auditor is unable to accept based on the test involving A″, then this strategy as described by Roberts (1978) calls for reassessment of both risks. The reassessed values of these risks are reflected in Figure 1 as TD_R and α_R.

Presumably, the failure to accept based on the analysis of evidence to this point would not lead to an increase in the acceptable level of either risk when this reassessment takes place; however, decreases in either may occur. A reduction of the risk of incorrect acceptance might be appropriate if, in the auditor's judgment, the sample evidence casts doubt on the appropriateness of the level of reliance on internal control used when initially assessing TD. Similar reassessments of that risk might be made because of changes in the perception of inherent risk, or the risk associated with other audit test results as compared to those used in the initial assessment of TD. The appropriate level for a revised risk of incorrect rejection might be lower than initially planned because a significantly larger than expected number of errors have been observed. The likelihood of encountering circumstances requiring adjustments may indicate that a reduction in the risk of incorrect rejection is warranted.

After reassessment of the two risks, this strategy, as described by Roberts, calls for a test of conclusiveness. The objective of such a test is to determine whether the sample evidence is sufficient to control both risks to their reassessed levels.[3] If the sample evidence is conclusive, an audit conclusion to reject the amount under examination is justified. Otherwise, the auditor will conclude that the sample evidence alone is insufficient for a final decision and some fallback option must be pursued. Generally these options may include: 1) expansion of the sample, where feasible; 2) performance of additional substantive procedures to provide additional evidence useful in fulfilling the audit objectives for which the statistical sample is germane; or 3) requesting that the client adjust or reconstruct the amount being examined.

Rejection Risk Control Strategy

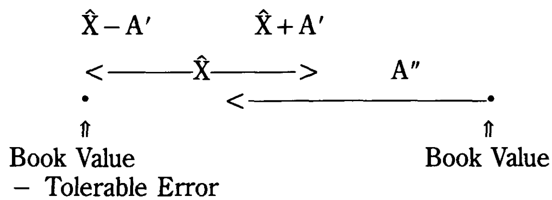
In contrast to the audit planning and evaluation strategy described above, statistical testing in many contexts other than auditing are based on direct control of the risk of incorrect rejection with the risk of incorrect acceptance not explicitly considered. As a result much nonauditing-statistical sampling literature is based on direct control of the risk of incorrect rejection. Therefore, many computer programs that may be useful for sample evaluation provide output based on control of the risk of incorrect rejection. The AICPA audit and accounting guide, *Audit Sampling,* considers sample determination and evaluation in this situation. In addition, sample evaluation strategies conceptually based on direct control of the risk of incorrect rejection and indirect control of the risk of incorrect acceptance can be found in the auditing literature [e.g. Arens and Loebbecke, 1981 and Bailey, 1981].

The sample size calculation described in *Audit Sampling* (pp. 93-94) permits control of the risks of incorrect acceptance and incorrect rejection to desired levels by varying the ratio of the desired allowance (A) to the tolerable error based on the table of ratios in Appendix C of the guide (p. 115). If the sample statistics are reported in the context of incorrect acceptance (i.e., A') or if the evaluation process is to focus on A', the audit guide discusses an evaluation strategy that may be used (pp. 94–99). This strategy is pictured in Figure 1 and is termed the "rejection risk control option." The first step in this process is to ensure that the *ex post* level of control of sampling risks can be at least equal to the planned level of control by determining that the condition $A' < A$ exists. If not, the sample is regarded as insufficient and fallback options must be considered. If $A' < A$, a direct test can be employed. If $B \in \hat{X} \pm A'$, the reported value can be accepted. In contrast to the acceptance risk control options, the rejection risk control option strategy initially tests with the risk of incorrect rejection set to the original planned level. In this case, the risk of incorrect acceptance is allowed to vary, and it will be at a lower level than planned (except in the rare case where $A' = A$, when it will be at the planned level). If $B \in \hat{X} \pm A'$ is not true, additional steps are suggested by the audit guide. They are described below. These steps ensure that the reported amount will not be rejected simply because sample estimators are more efficient than planned.

If $A' < A$ but B does not exist in the region $\hat{X} \pm A'$, the auditor may still be able to accept without computing an allowance for sampling risk related to the risk of incorrect acceptance. To do so, two conditions must be met. One condition requires that $\alpha < 2TD$. This condition ensures that an allowance for sampling risk based on α will also be associated with a risk of incorrect acceptance that is no more than TD. To ensure that such is the case, the reliability coefficient used in computing the (far) end of the range $\hat{X} \pm A'$ in relation to book value must be greater than the reliability coefficient which would be used in determining A'' and the associated allowance for sampling risk as related to TD. Because the reliability coefficient for α risk is associated with two-tail testing, that coefficient will be greater than the reliability coefficient for

TD only if the condition $\alpha < 2TD$ exists. Satisfying this condition effectively eliminates the possibility that $B-(\hat{X}-A')$ could be less than TE or that $(\hat{X}+A')-B$ could be less than TE even though B does *not* exist in $\hat{X} \pm A''$.

Without the first condition the situation which follows could arise, when $\alpha > 2TD$:



This would indicate acceptance even though the risk of the incorrect acceptance is greater than TD.

The second condition requires that $\hat{X} \pm A'$ exist in the region $B \pm TE$. In other words, it requires that the difference between the book value and the end of the range $\hat{X} \pm A'$ farther from the book value be less than the amount TE. When this condition is met, the risk of incorrectly rejecting the notion that the proper value of the account or balance in question is not materially different from the book value is less than (or equal to) the level α initially established for control of the risk of incorrect rejection.

If either of these two conditions fails to be met, computation of an allowance for sampling risk based on TD should be undertaken. Sample evaluation is then conducted in accordance with the acceptance risk control option steps previously discussed.

The Two Strategies Contrasted

As discussed, the acceptance risk control option will permit an initial acceptance test regardless of the relationship between A′ and A. If $A' < A$ then the sample size is sufficient to control the risks of incorrect acceptance and incorrect rejection to the planned levels. On the other hand, if $A' > A$, then the sample is not sufficient to control both risks to the planned level, and the initial decision process holds the risk of incorrect acceptance to the planned level by using the decision interval $\hat{X} \pm A''$. The auditor will not permanently reject the reported amount based on this decision process; however, if rejection were allowed, the risk of incorrect rejection would be greater than planned where $A' > A$.

The sufficiency test within the rejection risk control strategy prevents such an occurrence. This is accomplished by declaring the sample to be inclusive and then pursuing fall-back options in any instance for which $A' > A$. In all other instances $A' < A$.

Without the sufficiency test, classical statistical hypothesis evaluation using the acceptance risk control option is more likely to lead to acceptance than would the rejection risk control or sufficiency test options. Two conditions are necessarily associated with those sampling outcomes that lead to acceptance under the one strategy but not in the other. First, the *ex post* estimate of variability must exceed the level used in sample size determination. Second, the allowance for sample risk associated with *ex post* control of the risk of incorrect acceptance at the planned level must be small enough to warrant rejection of the alternative hypothesis (i.e., $B \epsilon \hat{X} \pm A''$).

It is possible to employ the acceptance risk control strategy with a sufficiency test by including the sufficiency test option with the acceptance risk control option as shown in Figure 1. In this case, the condition $A' < A$ would always exist under both approaches, and rejection could be held to levels equal to or less than originally planned under either strategy. However, the two strategies differ as to which risk to hold at the original level.

The acceptance risk control option with its decision rule based on the interval $B \epsilon \hat{X} \pm A''$ holds the risk of incorrect acceptance to the originally planned level. If we assume the sufficiency test, then the risk of incorrect rejection will be allowed to vary and will be smaller than originally planned.

By contrast the rejection risk control option with its decision rule based on the interval $B \epsilon \hat{X} \pm A'$ holds the risk of incorrect rejection at the planned level, and the risk of incorrect acceptance is allowed to vary and will be smaller than originally planned.

Thus both strategies will hold one risk at the planned level and allow the other risk to vary to a level lower than originally planned. The rationale for holding either risk at the planned level and allowing only the other risk to vary has not been adequately discussed in the literature. The rationale for either approach may appear to be questionable if we assume that the auditor considered, even in an intuitive way, the possible losses that might be associated with incorrect rejection or acceptance.

Both risks α and TD can be reduced by increasing sample size, but for any size sample α and TD have a wide range of trade-offs. These factors must be considered, at least intuitively, in deciding on planned levels of α, TD and sample size. Presumably the auditor balances the expected loss from each risk in some way when attempting to minimize the total expected loss from testing. The fact that the level of α and TD are often not the same may imply that the associated losses are also not equal. If so, it is not clear that either *ex post* strategy of holding one risk at the planned level will be optimal from an expected loss perspective.

By now it is clear that the choice of an appropriate evaluation strategy is less than obvious. A more formal examination of the implicit preferences employed when judging the sufficiency and competence of evidence using alternative strategies follows. An additional strategy is then developed. This additional strategy—the balancing strategy—seems logically defensible in relation to the formal analysis of the differences in extant strategies.

One means of more formal examination is to consider the expected value of the alternatives suggested by the alternative options. For simplicity we assume risk neutrality. In turn, we shall examine each of the two primary decision rules.

Within the context of the necessary conditions for different sample evaluation outcomes, the probability of incorrect acceptance is, of course, TD, if the acceptance risk control option is employed. If there is no error in the amount being tested then the probability of (correct) acceptance is the complement of the risk of incorrect rejection associated with A''. We designate this probability as $1 - \alpha''$. The numeric value of α'' may be determined after computing the associated reliability coefficient. The appropriate two-tailed reliability coefficient can be computed by determining the number of standard errors of the estimator contained in A″ ($A'' \div S_{\hat{X}}$).

Under the assumption that the auditor will act rationally to minimize maximum expected loss, acceptance using the primary rule from the acceptance risk control option implies that $\alpha''\ell_1 > TD\ell_2$, where ℓ_1 is the loss associated with incorrect rejection and ℓ_2 is the loss associated with incorrect acceptance. Under this assumption $\alpha''\ell_1$ is the maximum expected loss associated with a decision to reject the book value and $TD\ell_2$ is the maximum expected loss associated with a decision to accept. The relation, $\alpha''\ell_1 > TD\ell_2$, holds true without regard for the specific value of α'' since the value of α'' does not influence the decision rule calling for acceptance when $B \in \hat{X} \pm A''$. The rule is based on TD alone. In the extreme, this implies that even as the risk of incorrect rejection disappears ($\alpha'' \rightarrow 0$), or simply becomes very much smaller than TD, the consequences of incorrect rejection heavily outweigh the consequences of incorrect acceptance at the planned level. Such a conclusion requires that $\ell_1 >> \ell_2$, which is counter-intuitive. It demands that the negative consequences of incorrect rejection far exceed the negative consequences of incorrect acceptance. This seems a particularly undesirable artifact of any audit strategy since, in the extreme, it may favor accepting client results when the probability of their being correct is significantly smaller than the probability that they are without material error.

On the other hand, this decision rule seems to have greater intuitive appeal when $\alpha'' \rightarrow 1$ or whenever $\alpha'' >> TD$. In such circumstances the primary decision rule from the acceptance risk control option implies that $\alpha''\ell_1 < TD\ell_2$ and hence that $\ell_1 << \ell_2$. This result seems intuitively more appealing.

As a prima facia matter, this observation seems to favor the more liberal acceptance strategy associated with the acceptance risk control option. On those occasions when the other option employs this decision rule as a secondary criterion it is subject to the same criticism concerning the consequences of $\alpha'' \rightarrow 0$ or α'' becoming much smaller than TD because of unanticipated efficiency of the sampling process. On the other hand, because this option employs the adequacy criterion ($A' < A$) as a necessary condition for acceptance, it prevents the rule from operating and hence from indicating acceptance in those very circumstances where the rule seems intuitively most appealing. This occurs because α'' is less than the reliability coefficient for α. This can occur only when the allowance for sampling risk based on *ex post* control of TD at the planned level forces the range of estimators which leads to acceptance to be contained in a quite small region about the book value. Such limits on the range of acceptable estimators will approach the book value from above and below only as the variability of sampling results increases from planned levels. Of course this is the very condition which will cause the adequacy criterion test to nullify use of the decision rule by screening out the sample result as unacceptable.

Acceptance using the primary rule of the rejection risk control option requires exploration of the adequacy criterion. The adequacy criterion rule suspends judgment when $A' > A$. Suspension is called for regardless of the relationship between the projected audit value and the acceptance region about the book defined by controlling the risk of incorrect acceptance at the level TD.

Obviously, this suggests that the expected loss associated with (α, TD′) for all TD′ greater than TD exceeds some maximum acceptable level (where TD is the *ex post* probability of incorrect acceptance associated with the region $\hat{X} \pm A'$, which controls for α at the planned level). In addition the sufficiency test suggests that the maximum acceptable expected loss associated with reliance on sample evidence should be $\alpha \ell_1 + TD\ell_2$. If this condition cannot be achieved based on results of a sample, then incurring the costs associated with fallback option(s) becomes necessary. Any such fallback should be planned to produce sufficient additional competent evidential matter. Theoretically, planned fallback should reduce the risk of incorrect acceptance to the level TD while maintaining α at the planned level.

Conversely, the acceptance risk control option, without the sufficiency test as a primary screen, may permit acceptance without regard to the implicit level of the risk of incorrect rejection associated with its primary test which is based on TD alone. As pointed out above, this may implicitly allow α'' to become quite large when the variability of sample results exceeds planned levels. Therefore it might be inferred that $\alpha'' \ell_1 + TD\ell_2$ is small enough to negate the cost benefit of fallback procedures even when $\alpha'' \rightarrow 1$. This seems an undesirable result. It suggests that *either* available fallback options are 1) extremely costly, 2) inefficient, or 3) ineffective at reducing risk of incorrect rejection (e.g. $\alpha'' \ell_1 < \alpha \ell_1$ + cost of employing feasible fallback option(s)); *or* that the loss associated with incorrect rejection is trivial ($\ell_1 \approx 0$).

If the latter were true, there would be no reason to have controlled incorrect rejection risk in the first place during sample size planning. If something from the former set of conclusions is true then no cost effective practical means for further reducing risk is available after sampling *nor* were such procedures considered subsequently available prior to sampling. Had they been considered subsequently available then the risk of incorrect rejection would have been worth controlling explicitly in formulation of the primary decision rule.

These results seem to favor use of the acceptance strategy associated with the rejection risk control option rather than the more liberal acceptance strategy of the acceptance risk control option. Of course, this finding is in direct conflict with the prior prima facia results which favored the logic of the acceptance risk control option. This paradox suggests that another strategy for sample evaluation be contemplated.

The Balancing Strategy

As depicted in Figure 2, the balancing strategy begins with and employs the same straightforward decision rules as the other strategies when $A' = A$. When $A' \neq A$, the adequacy criterion rule (as employed by the rejection risk control and sufficiency test options) is invoked as a primary screen. When results indicate that the variability of sample observation exceeds the planned level ($A' > A$), the sample is deemed inconclusive and appropriate fall-back options are considered. This is also consistent with the other adequacy criterion options.

The balancing strategy takes its unique character from its next stage decision rule. When invoked, the rule calls for acceptance of the amount being tested if the book value falls in the region $\hat{X} \pm A^b$, where A^b is the monetary amount which balances the *ex post* risks of incorrect rejection, α_b, and incorrect acceptance TD_b, such that $\alpha/TD = \alpha_b/TD_b = \ell_2/\ell_1$. This condition is equivalent to $\alpha_b\ell_1 = TD_b\ell_2$. When the expected losses of incorrect rejection and incorrect acceptance balance one another in this fashion, the critical limits based on control of α_b and TD_b respectively will be equivalent. In each case these limits are $\hat{X} \pm A^b$. The determination of A^b requires simultaneous solution of the following equations:

$$2F_N(Z_{\alpha_b/2})/F_N(Z_{TD_b}) = C_1$$

$$TE/S_{\hat{X}} = Z_{\alpha_b/2} + Z_{TD_b} = C_2$$

where $C_1 = \alpha/TD = \ell_2/\ell_1$ and C_2 = the number of standard deviations of the sampling distribution in the region bounded by the null and alternative hypotheses. $F_N(\bullet)$ is the cumulative standard normal density function for the specified standard deviate. $Z_{\alpha_b}/2$ is the number of standard deviates which provide for control of the risk of incorrect rejection at level α_b and Z_{TD_b} is the number of standard deviates which provide for control of the risk of incorrect acceptance at level TD_b.

There is no closed formed analytical solution to these two equations because the $F_N(\bullet)$'s are integrals of a normal probability function. However, as a practical matter, numerical approximate functions (e.g., see Abramowitz and Stegun, 1964, p. 299) can readily be employed to produce $F_N(\bullet)$ values. Other numerical algorithms may be used in conjunction with these approximations to compute $Z_{\alpha_b/2}$. Once a solution for $Z_{\alpha_b/2}$ is computed then $A^b = Z_{\alpha_b{}^2/2}\, S_{\hat{X}}$ is available and the decision rule can be employed based on whether $B \in \hat{X} \pm A^b$.

By employing both the sufficiency test rule and balancing rule, the balancing strategy avoids the pitfalls associated with prior strategies. The sufficiency test, as a primary rule, assures that consideration of fall-back procedures will not be ignored and that the consequences of incorrect rejection will not be treated as trivial. In this sense, it is equivalent to the rejection risk control option and sufficiency test option which dominate the acceptance risk control option with respect to primary rule selection.

If $A' < A$, the balancing rule, when allowed to operate, will reduce both risks below planned levels. Therefore, $A'' > A^b > A'$. Acceptance will occur less frequently with the balancing strategy than either of the other strategies. The balancing strategy has a higher potential for failing to accept than the acceptance risk control option because it employs the sufficiency screen and because the critical acceptance region for secondary testing is smaller, $\hat{X} \pm A^b$, than the region of acceptance, $\hat{X} \pm A''$, associated with the acceptance risk control option. It is also more conservative than the other sufficiency test options with which it shares the primary rule because they also rely for secondary testing on the larger region $\hat{X} \pm A''$.

The more efficient the sample in relation to planned efficiency the closer A^b will be to the midpoint between the alternative hypotheses. (When $\alpha = 2TD$, $A^b = A$ without regard to sample size because the reliability coefficients for both risks will be equal, before and after sample results are available.) For $\alpha < 2TD$, A^b and hence the acceptance region $\hat{X} \pm A^b$ will become smaller as sampling efficiency improves. For $\alpha > 2TD$ the acceptance region $\hat{X} \pm A^b$ becomes larger as sampling efficiency increases.

By converging on the midpoints between hypotheses as critical limits, the rule assures that as α_b approaches 0 so too will TD_b (and vice versa), thus permitting the expected loss from either error to be reduced from $\alpha\ell_1 + TD\ell_2$ to $\alpha_b\ell_1 + TD_b\ell_2$ with $\alpha_b < \alpha$ and $TD_b < TD$, while maintaining control of both risks.

The balancing strategy concludes with the same decision rules as the other strategies, except that the balancing strategy rebalances A^b, in accordance with the ratio of α_R/TD_R when considering the adequacy of adjustments in relation to statistical results.

Other Issues

The previous sections have been concerned with a single issue—the merits of alternative strategies that are available to the auditor when the *ex post* efficiency of the statistical estimator appears to be different from the planned level of efficiency. More specifically, what is the nature and result of the trade-offs between the risks of incorrect acceptance and incorrect rejection that are implied by several commonly available alternatives? In addition, we considered a possible strategy for determining levels of these risks by incorporating the losses that might be associated with these risks.

We also briefly consider some other unresolved issues in classical auditing sampling in the following sections. These issues have only recently been recognized by researchers in the audit sampling literature, and may prove to be fertile ground for future research.

Assessing the Risk of Incorrect Acceptance

A good deal of work has been produced suggesting that the assessment of TD is a tricky task and that current models of determining that risk level for sample evaluation purposes are overly simplistic. Both Leslie [1984] and Kinney [1984] and implicitly the CICA study, *Extent of Audit Testing* [1980], point out that the current SAS 47 approach for developing TD may be viewed as intending TD to be a conditional risk. Under this view, the SAS 47 approach invokes TD as a conditional posterior risk. This is the risk, *given that material error exists,* that the auditor will incorrectly accept. This may be significantly less than the Bayesian type posterior risk of incorrect acceptance which would consider the conditional probability for incorrect acceptance in relation to the marginal probability of acceptance, where the marginal is the probability of sample results leading to acceptance without regard to whether that decision to accept is correct. Among other problems associated with risk assessment are the need to contend with the impact of artificial specification of simple rather than complex hypotheses [Dacey & Ward, 1986] and the potential benefit of considering extension of Bayesian type models to include posterior consideration of correct and incorrect acceptance in relation to the results of all evidential procedures rather than only detailed sampling procedures. In addition, as highlighted by Cushing and Loebbecke [1983], nonsampling risks may not be as limited in their potential impact as current practices would suggest.

Ex Post Sampling Risk

Beck and Solomon [1985] have observed that the achieved sampling risks may be dependent upon the decision rule used and the estimator selected when highly skewed populations force defacto violation of the normality assumptions associated with the sampling distribution. This observation suggests that the auditor faces different *ex post* risks exposures and hence different audit consequences when the statistical assumptions are violated. Under such conditions, it becomes important for the auditor to choose an appropriate estimator and an appropriate decision rule for evaluating the sample results so that he can minimize his risks exposure. The Beck and Solomon study provides suggestions for meeting this objective by pairing decision rules with statistical estimators based upon an *ex post* analysis of the sample evidence (e.g., error pattern).

The (two) decision rules that Beck and Solomon refer to are based on the two alternative hypothesis testing approaches. Under one approach the auditor tests null hypothesis that the account book value is fairly presented (the decision rule based on this approach has been referred to as Elliott and Roger (E & R) decision rule). In essence, this is a test of the type associated with the rejection risk control option described above. Under the second approach the null hypothesis being tested is that the account book value is misstated by an amount greater than tolerable error. This approach was used in Statement on Auditing Procedure (SAP) 54. This is a test of the type associated with the primary decision rule from the acceptance risk control option as discussed above. It should be mentioned here that the E & R and SAP 54 decision rules are equivalent for planning purposes as demonstrated by Roberts [1974] when normality of the sampling distribution is assumed.

Beck and Solomon then illustrate how the achieved sampling risks are changed when the decision rule used is changed. Assume that the accounting population is highly skewed (as is often the case usually, see Stringer, 1963) to the right and the estimator used is the ordinary mean per unit (MPU) estimator. Since the accounting population is highly skewed, the MPU estimates are likely to exhibit skewness, and in the presence of skewness the estimator of the population mean and the estimator of the standard error are found to be highly positively correlated (see Neter and Loebbecke, 1975). Suppose now that the client's asset account book value is fairly stated, but the auditor's sample estimate of the account mean (total) value is drawn from the lower region of the sampling distribution and thus is less than the actual mean (total) value of the account. Since the estimator of the mean is positively correlated with the estimator of the standard error, a smaller than average mean estimate would be accompanied by a smaller than average standard error estimate. In this situation, the two-sided confidence intervals computed under the E & R decision rule would be centered below the actual mean and also would be too narrow. Consequently, the risk of efficiency error would be higher than what was planned. However, when the SAP 54 decision rule is used, because of small estimates of the mean and standard error, a large estimate of monetary error would result and with a smaller achieved precision measure the risk of efficiency errors would become smaller than the risk determined using E & R decision rule. A similar argument can be presented for the risk of effectiveness which also is lower under the SAP 54 decision rule than under the E & R decision rule when the mean estimate is such that a larger than average estimate of standard error is projected from sample results.

Asymmetric Materiality Thresholds

There is empirical evidence suggesting that decisions about materiality may not be symmetric. In some circumstances auditors may be less tolerant of overstatement than understatement and wish to establish audit testing hypotheses accordingly (Ward, 1976). Recently, Srivastava and Ward [1986] have developed a methodology that incorporates such an asymmetry for variable sampling. Their preliminary results show that the auditor can achieve a significant reduction in the sample size when the asymmetric materiality thresholds are used in the planning stage. It is interesting to note that the sample size reduction is achieved without sacrificing the two-tail test for control of the risk of incorrect rejection.

Conclusion

The objective of this paper was to identify and discuss some unresolved issues in classical audit sample evaluations. The selection of which issues to consider was not random and, in fact, was very biased. The bulk of the paper was devoted to a discussion of the implications of common evaluation strategies that are presented in the audit sampling literature for situations where the achieved efficiency of the estimator appears to be different from the planned efficiency. When this occurs, both the acceptance risk control and the rejection risk control strategies create a decision interval such that one risk (TD or α) is held to the originally planned level and the other risk is allowed to vary from the planned level. Little discussion is presented in the literature concerning the rationale for selection of one or the other risk to hold at the planned level, or why it is so logical to allow the other to vary from the planned level. In fact, this type of trade-off process may seem contradictory if there is at least a rough, intuitive balancing of expected losses from the two risks when the acceptable risk levels are initially planned. From this viewpoint, a strategy was presented which attempts to balance the expected losses for the two risks based on *ex post* information. This process would appear to have some conceptual merit and to warrant further investigation. In addition, brief comments on several other recently discussed issues were presented. Although these issues have just been identified and thus are perhaps further from solution, they merit mention and probably future discussion and investigation.

End Notes

1. Much of the discussion of evaluation of samples in Bailey is based on the same premise as the audit guide option approach; however, he recognizes the alternative approach similar to the acceptance risk control strategy in footnotes.

2. Guidance for establishing the two risk levels, TD and α, is available elsewhere. See, for example, Arens & Loebbecke [1981, p. 136] and SAS 39. A significant amount of prior effort has been expended to assist the auditor in understanding how to establish an appropriate level for TD. Some issues and problems raised by these studies are reviewed in a separate section of this paper.

3. The statistical evidence may be considered conclusive if the number of standard errors of the estimator contained in the tolerable error amount, TE, for the account being tested exceeds the sum of the number of standard errors of the estimators required to control α_R and TD_R at the reassessed levels.

References

Abramowitz, M. & I.A. Stegun, *Handbook of Mathematical Functions,* N.Y.: Dover Publications, Inc. (1964).

American Institute of Certified Public Accountants, *Audit and Accounting Guide: Audit Sampling,* New York (1983).

Arens, A.A. and J.K. Loebbecke, *Applications of Statistical Sampling to Auditing,* Englewood Cliffs, N.J.: Prentice-Hall (1981).

Auditing Standards Board, *Statement on Auditing Standards,* No. 39, N.Y.: AICPA (1981), and *SAS No. 47,* N.Y.: AICPA (1983).

Bailey, A.D., Jr., *Statistical Auditing: Review, Concepts, and Problems,* N.Y.: Harcourt Brace Jovanovich, Inc. (1981).

Beck, P.J. and I. Solomon, "*Ex Post* Sampling Risks and Decision Rule Choice in Substantive Testing," *Auditing: A Journal of Practice and Theory* (Spring, 1985) pp. 1-10.

Blocher, E. and J.H. Bylinski, "The Influence of Sample Characteristics in Sample Evaluation," *Auditing: A Journal of Practice and Theory* (Fall, 1985) pp. 79-80.

Cushing, B.E. and J.K. Loebbecke, "Analytical Approaches to Audit Risk: A Survey and Analysis," *Auditing: A Journal of Practice and Theory* (Fall, 1983) pp. 23-41.

Dacey, R. and B. Ward, "On the Fundamental Nature of Professional Opinions: The Traditional, Bayesian and Epistemic Methods of Attestation," *The Sixth Illinois Auditing Research Symposium,* Champaign, IL: University of Illinois Department of Accountancy (1986), forthcoming.

Ernst & Whinney, *Auditing Sampling: College Edition,* Cleveland: E & W (1979).

Felix, W.L., Jr., D.A. Leslie, and J. Neter, "University of Georgia Center for Audit Research Monetary-Unit Sampling Conferences, March 24, 1981," *Auditing: A Journal of Practice and Theory* (Winter, 1982) pp. 92-103.

Guy, D.M. and D.R. Carmichael, *Auditing Sampling: An Introduction to Statistical Sampling in Auditing,* N.Y.: John Wiley & Sons (1986).

Kinney, W.R., Jr., "Discussant's Response to an Analysis of the Audit Framework Focusing on Inherent Risk and the Role of Statistical Sampling in Compliance Testing," *Auditing Symposium VII,* Proceedings of the 1984 Touche Ross/University of Kansas Symposium on Auditing Problems, Lawrence, KS: University of Kansas School of Business (1984).

Leamer, E.E., *Specification Searches: Ad Hoc Inference with Nonexperimental Data,* N.Y.: John Wiley and Sons (1978).

Leslie, D.A., "An Analysis of the Audit Framework Focusing on Inherent Risk and the Role of Statistical Sampling in Compliance Testing," *Auditing Symposium VII,* Proceedings of the 1984 Touche Ross/University of Kansas Symposium on Auditing Problems, Lawrence, KS: University of Kansas School of Business (1984).

Neter, J. and J.K. Loebbecke, *Behavior and Major Statistical Estimates in Sampling Accounting Population,* N.Y.: AICPA (1975).

Roberts, D.M., *Statistical Auditing,* N.Y.: AICPA (1978).

Srivastava, R.P. and B. Ward, "An Investigation of Sampling Efficiency in the Presence of Asymmetric Materiality Threshold," (Unpublished working paper).

Stringer, K.W., "Practical Aspects of Statistical Sampling in Auditing," *Proceedings of the Business Economics and Statistics Section,* American Statistical Association (December 1963).

Study Group, *Extent of Audit Testing: A Research Study,* Toronto: Canada Institute of Chartered Accountants (1980).

Thompson, J.H., *Establishing Alpha and Beta Risk Levels in the Evaluation of Substantive Statistical Auditing Tests that Employ Hypotheses Testing,* Ph.D. Dissertation, The University of Oklahoma (1982).

Ward, B., "An Investigation of the Materiality Construct in Auditing," *Journal of Accounting Research* (Spring, 1976) pp. 138–152.