Symposium on Auditing Research IV

By an Audit Group at the
University of Illinois at Urbana-Champaign
Department of Accountancy
Symposium on Auditing Research IV

By an Audit Group at the University of Illinois at Urbana-Champaign

Department of Accountancy
University of Illinois at Urbana-Champaign
PREFACE

The papers contained in this volume were presented at the Fourth Symposium on Auditing Research held at the University of Illinois at Urbana-Champaign on November 13-14, 1980. The purpose of the symposium was to provide a forum for the discussion of current research on auditing issues of interest to both academicians and practitioners.

Eight research papers were presented at the symposium. For each paper, there was at least one discussant. The papers were presented by academicians; the discussants came from both the practice and academic sides of the profession. The papers cover a variety of contemporary auditing topics. The papers, together with the comments, are intended to provide a useful resource for those interested in the subject matter. We also hope that the interchange of ideas which this forum provided will stimulate further research into these and other issues.

An Audit Group at the University of Illinois at Urbana-Champaign.

Richard E. Ziegler
Clifton E. Brown
Philip E. Fess
Frederick L. Neumann
Soong H. Park
Joseph J. Schultz, Jr.
FOREWORD

This is the fourth in a series of Audit Symposia sponsored by the Department of Accountancy at the University of Illinois at Urbana-Champaign. Originated at the suggestion of Professor Philip E. Fess, it has, over the years, become a major forum for the discussion of research related to auditing.

The variety of topics encompassed by any one Symposium has increased as attention given this significant professional responsibility has grown and broadened in scope. The accelerating pace and complexity of modern economic life have enhanced the demand for swift and reliable information. Auditing has earned a respected role in this process by constantly striving to improve its performance.

Other influences have also contributed to the expanded considerations of auditing as a focus of research. The recent special attention given to accountability has served to augment another dimension of the auditor's responsibility and has prompted expansion of research in the area of control. Technical developments in quantitative methods have led to extended use of statistics. Increasingly, research in this area has originated because of audit considerations instead of being derived solely from existing knowledge in other areas. Greater emphasis on the human element has brought recognition that neither auditing nor its products are neutral in their effects.

We are pleased that these Symposia have made a contribution to research in auditing. In addition to the various papers which have reported on investigations into these several areas, the
Symposia have provided opportunity for both formal and informal discussion. The blend of practitioners and scholars among the participants has made these sessions particularly valuable as forums for addressing these issues. The continued respect for and reliance upon the auditing function by society demands an increasing and imaginative assault on the unsolved problems of the discipline, through research.

We are indebted to and wish to acknowledge gratefully the financial assistance of Arthur Andersen & Co. This generous support made possible this Fourth Symposia and the publication of the papers presented.

Frederick L. Neumann, Head
Department of Accountancy

Vernon K. Zimmerman, Dean
College of Commerce and Business Administration
Decision - Theoretic Estimation Methods
in Accounting and Auditing

by

Barry E. Cushing

Graduate School of Business
The University of Utah

Prepared for Symposium on Auditing Research IV, University of Illinois,
November 13-14, 1980.
Decision - Theoretic Estimation Methods
in Accounting and Auditing

The objectives of this paper are (1) to develop a decision theoretic model which will identify optimal sample sizes and terminal point estimates for estimation problems often faced by accountants and auditors, and (2) to test the sensitivity of the minimum expected total costs projected by this model to deviations from prescribed model solutions and to mis specification of model inputs. While it is hoped that the model developed here may be usefully employed in certain circumstances, the more fundamental contribution of the paper lies in the insight to sample planning issues provided by the sensitivity results.

The problem of estimating the value of an accounting variable is frequently encountered by accountants and auditors. For example, Newman [1976, pp. 5-8], Akresh [1978], and Petti [1978] identify a wide variety of accounting estimation problems drawn from their experience. These include determination of physical inventories, valuation of loans and receivables, establishment of reserves, assessment of the impact of using an alternative accounting method, determination of interairline billings, measuring the current portion of installment receivables, and many others.

In certain cases, as described by Meikle [1972, p. 14], Leslie [1977], and Roberts [1978, pp. 39-40], the auditor faces estimation problems similar or identical to those of the accountant. Furthermore, while hypothesis testing is generally regarded as the appropriate paradigm for the attest function [Elliott and Rogers, 1972], rejection of the hypothesis that an account's book value is fairly stated leaves the auditor with the problem of estimating a reported value for that account. Finally, it could even be argued that hypothesis
testing is merely a special case of estimation in which a pre-established
decision rule is imposed upon the estimation result. I do not wish to debate
this last point here, but I do contend that the estimation of accounting values
is an issue which should be of general concern to auditors as well as to
accountants.

One common denominator of many of the estimation problems cited above is
that the accountant or auditor possesses a significant amount of prior infor-
mation concerning the quantity to be estimated. For example, he may have
access to detailed records of a closely related variable, such as the book
value of receivables, inventories, or fixed assets. Furthermore, he is
usually familiar with the information system which produced those records,
and often has had the experience of estimating the quantity during previous
accounting periods. A topic frequently addressed in contemporary auditing
literature is the issue of how to incorporate this prior information into
the planning of substantive tests. One possible approach is to incorporate
the prior information available to the accountant or auditor into a prior
probability distribution, and then use sample results to revise this distribu-
tion. The resulting posterior distribution could provide a more economical
and systematic way of considering all available sources of information in
achieving the estimation (or decision) objective. Furthermore, if the accountant
or auditor can express the consequences of estimation error in the form of a
loss function, a formal decision theoretic model may be used to determine the
extent of sampling which is optimal. This paper develops a model which may
be used in this manner.

In Part I of the paper, the elements of the decision theoretic model are
developed and discussed. Following this, Part II explains the sensitivity
tests on the model and examines their results. Part III is a brief summary
of the paper's major findings and conclusions.

I. THE MODEL

Decision theoretic models previously developed by Schlaifer [1959] and Raiffa and Schlaifer [1961], have been proposed for application in accounting and auditing settings by Scott [1973] and Moriarty [1973]. However, Scott defines the sampling unit as the entire set of transactions for one day; as pointed out by Felix and Grimlund [1977, p. 26], this assumption renders the Scott approach of limited practical usefulness. Moriarty assumes a likelihood function defined by the mean-per-unit estimator using simple random sampling; as demonstrated by Neter and Loebbecke [1975, pp. 36-39], this estimator is extremely imprecise for developing accounting and auditing estimates, and furthermore is not entirely reliable for highly skewed populations. Moriarty also uses the quadratic loss function throughout his work, and as this paper will show, such an assumption severely limits the generality of his conclusions.

Other decision-theoretic approaches to sampling in accounting and auditing have been proposed by Kaplan [1973], Kinney [1975a, 1975b], Kinney and Warren [1979], and Felix and Grimlund [1977]. The Kaplan paper proposes in very general terms a Bayesian extension of classical estimation methods, but does not examine the more specific issues of sample size calculation and incorporation of the posterior distribution and loss function into the decision process of the accountant or auditor. Kinney [1975a] develops a decision theoretic variables sampling model which is an extension of the classical hypothesis testing approach advocated for audit sampling by Elliott and Rogers [1972]; further extensions of this approach are developed by Kinney [1975b] and Kinney and Warren. Aside from the fact that its primary objective is audit testing rather than accounting estimation, Kinney's model also is based upon the imprecise
and unreliable simple mean-per-unit sampling technique. Furthermore, Kinney also assumes only a single form of loss function, namely a step loss function, which limits the generality of his conclusions.

Felix and Grimmelund develop a Bayesian procedure for estimating the total error in an account balance. Their model also draws on the work of Raiffa and Schlaifer [1961]. However, their model deals primarily with techniques for generating a posterior distribution, and does not incorporate loss functions or optimal sample size calculations.

The model proposed in this paper is a variation of the preposterior decision analysis developed by Raiffa and Schlaifer [1961, pp. 188-205]. The estimation technique chosen is the stratified mean-per-unit estimator, and accordingly the sampling distribution is assumed to be normal in form. In this situation the natural conjugate prior is also normal, and therefore it is assumed that the prior beliefs of the accountant or auditor can be represented, at least approximately, by a normal distribution. The sampling cost function is assumed to be linear and include both a fixed and variable component. Finally, rather than select a specific form of loss function, I have chosen to develop four different variations of the model, the first three of which incorporate linear, quadratic, and step loss functions, respectively, while the fourth utilizes a "target interval width" concept in lieu of a loss function to determine sample size. In the sections which follow, the notation associated with these various elements of the model is presented, and the assumptions outlined above are discussed. In the final section of Part I, the analysis necessary to determine sample sizes and optimal terminal point estimates is presented.

The Stratified Mean-per-Unit Estimator

The stratified mean-per-unit estimator was chosen as the basis for the
model developed in this paper for essentially three reasons, (1) it is
commonly used in accounting and auditing, (2) it performed quite favorably
with respect to both precision and reliability in the comprehensive empirical
study of accounting estimators by Neter and Loebbecke [1975], and (3) its
integration into a decision theoretic or Bayesian framework has not yet
appeared in the accounting literature, whereas such integrating models have
appeared for both the simple mean-per-unit estimator and dollar unit sampling. ¹

The notation used in this paper for the stratified mean-per-unit esti-
mator is adopted from Cochran [1963, pp. 88-97]. Basic parameters include
the following:

\[ N \quad \text{number of items in the population} \]
\[ L \quad \text{number of strata} \]
\[ h \quad \text{stratum index} \]
\[ N_h \quad \text{number of population items in hth stratum} \]
\[ w_h = N_h/N \quad \text{stratum weighting factor} \]
\[ n \quad \text{sample size} \]
\[ n_h \quad \text{number of sample items from hth stratum} \]
\[ y_{hi} \quad \text{value of the ith item from the hth stratum} \]
\[ S_h^2 \quad \text{true variance of the items in the hth stratum} \]

Once a sample has been drawn, the sample mean for the hth stratum is given by

\[
\overline{y}_h = \frac{\sum_{i=1}^{n_h} y_{hi}}{n_h}
\] ¹

¹Several models integrating Bayesian and decision theoretic approaches
with the simple mean-per-unit estimator have been previously cited. Models
integrating the Bayesian approach with dollar unit sampling have been presented
by Garstka [1977] and Vanecek [1978].
and an unbiased estimate of the population mean-per-unit is given by:

\[
\overline{y}_{st} = \frac{1}{L} \sum_{h=1}^{L} w_h \overline{y}_h
\]  

(2)

The sample data may also be used to develop an unbiased estimate of \(S_h^2\) for each stratum:

\[
s_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (y_{hi} - \overline{y}_h)^2
\]  

(3)

Define \(V(\overline{y}_{st})\) as the variance of \(\overline{y}_{st}\) as an estimate of the population mean. It can be shown that \(V(\overline{y}_{st})\) is minimized for a fixed sample of size \(n\) if:

\[
n_h = n \frac{W_h S_h}{\sum_h W_h S_h}
\]  

(4)

Under this minimum variance, or Neyman, allocation, \(V(\overline{y}_{st})\) is given by:

\[
V(\overline{y}_{st}) = \frac{(\sum_h W_h S_h)^2}{n} - \frac{\sum_h W_h S_h^2}{N}
\]  

(5)

where the second term on the right represents the finite population correction factor.

A key assumption underlying the use of the stratified mean-per-unit estimator in this paper is that its sample means are normally distributed for samples of the size typically drawn from accounting populations. This assumption has been termed "questionable" by Teitlebaum and Robinson [1975, p. 77], and therefore it is appropriate to briefly review the available evidence supporting the assumption. The empirical study by Neter and Loebbecke [1975] provides quite convincing evidence.
Neter and Loebbecke determined the proportion of confidence intervals constructed from 600 samples which contained the true population mean. Although they tested four different confidence intervals, the focus here is on the two-sided, 95.4% interval, which is very close to the 95% interval often recommended for accounting and auditing purposes [Newman, 1976, pp. 73-74; Elliott and Rogers, 1972, pp. 53-54]. To test the stratified mean-per-unit estimator, Neter and Loebbecke used (1) four different populations, (2) samples of 15 and 20 strata, (3) sample sizes of 100 and 200 items, and (4) population error rates of .5, 1, 5, 10, and 30%. This provided a total of 80 different realizations of the statistic "proportion of correct intervals" or 80x600 = 48,000 intervals in all.

Out of the total of 48,000 intervals, tabulations based upon data provided in their Tables A-71 through A-86 [pp. 188-195] show that 45,723, or 95.3% contained the population mean. This is very close to the theoretical percentage of 95.4%. Furthermore, since 95.4% of 600 replications is equal to 572.4, half of the 80 tests should have resulted in a number of correct intervals equal to 572 or less, and half in 573 or more. In reality 36 of the 80 tests resulted in 572 or fewer correct intervals, and 44 resulted in 573 or more correct intervals. Finally, according to Neter and Loebbecke [p.8], the estimated standard error of the proportion of correct confidence intervals in 600 replications, for a proportion of correct intervals of 95%, is equal to .009. Based upon this number, it may be predicted that 95% of their proportions of correct intervals, or 76 of 80, should have fallen within the range 95.4% ± 1.96 x 0.9%. In reality, 73 of 80, or 91.3% fell within this range. All of these results provide persuasive evidence in support of the assumption that the sampling distribution of the stratified mean-per-unit estimator is normally distributed in accounting and auditing contexts.
Another assumption inherent in the application of the stratified meanper-unit estimator in accounting and auditing is that the sample standard deviation, \( s_h \), provides an acceptable approximation of the population standard deviation, \( S_h \). Teitlebaum and Robinson also question this assumption. However, the technique of using \( s_h \) to approximate \( S_h \) was employed by Neter and Loebbecke in constructing their confidence intervals, and it is clear from the above review of their results that this did not cause significant reliability problems. In more general terms, it should be pointed out that from the accountant's or auditor's point of view the seriousness of an error in estimating the population standard deviation is an economic question rather than a statistical one. Both the accountant and the auditor are actors in an economic world, and as such may rationally tolerate minor violations of statistical assumptions as long as the economic consequences of doing so are not significant. Therefore, in order to fully evaluate Teitlebaum and Robinson's criticism, information is required concerning the economic consequences of incorrectly estimating the population standard deviation. Such information is provided by the sensitivity analysis results reported later in this paper.

**The Prior Distribution**

Refer to the true population mean-per-unit as \( \mu \). We require a prior probability distribution \( f'(\mu) \), or equivalently \( f'(M) \) where \( M \) is the true population total, or \( N\mu \). As previously mentioned it is assumed that the accountant's or auditor's prior probability distribution is normally distributed or can be closely approximated by a normal distribution. According to this assumption, \( f'(\mu) \) is given by:

\[
f'(\mu) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\mu - \mu')^2}{2\sigma^2}} \tag{6}
\]
where \( m' \) and \( \sigma' \) are the per unit mean and standard deviation, respectively, of the prior distribution.

Evidence on the nature of accountant's or auditor's subjective prior probability distributions over accounting variables is limited. The author is aware of two studies in which such distributions were assessed using the fractiles method with practicing auditors as subjects [Romney, 1977; Solomon, Krogstad, Tomassini and Romney, 1980]. These studies at least indicate that the subjects used were consistently willing and able to provide their prior distributions. However, in both studies there was no consistent pattern to the shape of the priors across subjects.

Even though accountant's and auditor's prior distributions for accounting variables may not be consistently normal, it may be possible to approximate them using a normal distribution without a material loss of accuracy. In order to evaluate this possibility, Cushing and Romney [1980] develop a least squares approximation technique to fit a normal distribution to each of the 192 prior distributions assessed by Romney [1977]. In terms of percentage deviations of the normal distribution fractiles from their counterparts in the assessed distributions, they found that the normal approximations were very close.

Of greater relevance than the accuracy of the normal approximation to the prior distribution are the issues of (1) the accuracy of the posterior distribution when a normal approximation to a non-normal prior is used, and (2) the opportunity loss arising from using a normal approximation to a non-normal prior. Evidence exists on both of these issues. With respect to the first, Schlaifer [1959, p. 448] concludes that:

If the variance of the decision maker's true prior distribution is large compared with the sampling variance of \( \overline{x} \), he can simplify his calculations with no material loss of accuracy by substituting the mean and variance of his true prior distribution into the formulas which apply to a Normal prior distribution.
Empirical data, described more fully in the section dealing with sensitivity analysis, indicates that the condition cited by Schlaifer will generally hold in an accounting context.

Evidence concerning the opportunity loss issue is provided both by Raiffa and Schlaifer [pp. 131-138] and Cushing and Romney. Raiffa and Schlaifer detail three examples in which a preposterior analysis is performed using a normal approximation to a radically non-normal beta prior distribution. Opportunity loss, representing the difference between the expected net gain of the true optimum sample size and the expected net gain of the sample size obtained using the normal approximation, was equal to 2.4%, 0%, and 0.9% of the optimum net gain in the three cases. Cushing and Romney construct ten audit decision problems incorporating the most nonnormal of the 192 subjective prior distributions assessed by Romney, and find that the opportunity loss from using the normal approximation averages only 3.6% of the minimum expected total cost obtained by using the true distribution.

In some cases the accountant or auditor who is unwilling or unable to specify a subjective prior distribution over an accounting variable may be able to establish an objective prior using regression analysis, as described by Stringer [1975] and by Gurry and Santi [1975]. In this case the prior will initially be in the form of a t distribution. If the number of observations used in the regression is large, a very close normal approximation may be obtained by substituting the standard deviation of the t distribution for $\sigma'$ and of course the mean of the t distribution for $m'$. When the number of observations is small, this approximation will be unsatisfactory, but a fitting technique such as that suggested by Cushing and Romney should give a very good normal approximation.

If a subjective prior distribution and objective prior data are available,
a combined prior may be obtained by treating the objective data as a sample result and using it to revise the subjective prior. The resulting posterior would then serve as a prior distribution for the stratified sample.

In light of the factors discussed above, it is concluded that the assumption of normality of accountant's and auditor's prior distributions for accounting variables is reasonable, at least in many cases. Given the parameters \( m' \) and \( s' \) for the normal prior distribution, and the sample results \( \bar{y}_{st} \) and \( V(\bar{y}_{st}) \), the next issue is the determination of the posterior distribution parameters \( m'' \) and \( s'' \). Paraphrasing Schlaifer [p. 441], it may be shown by calculus\(^2\) that if (1) the prior distribution of \( \mu \) is normal with parameters \( m' \) and \( s'^2 \), (2) the sampling distribution of \( \bar{y}_{st} \) is normal with parameters \( \mu \) and \( V(\bar{y}_{st}) \), and (3) the value of \( V(\bar{y}_{st}) \) is known, or if a reasonably reliable estimate is used, then the posterior variance is given by:

\[
\frac{1}{s''^2} = \frac{1}{s'^2} + \frac{1}{V(\bar{y}_{st})} \tag{7}
\]

and the posterior mean by:

\[
m'' = \frac{(1/s'^2)m' + (1/V(\bar{y}_{st}))\bar{y}_{st}}{(1/s'^2) + (1/V(\bar{y}_{st}))} \tag{8}
\]

The Loss Function

The loss function reflects the economic consequences of differences between the accountant's or auditor's terminal point estimate \( R \) and the true population total \( M \). A linear loss function is given by:

\[\text{Loss} = (R - M)^2\]

\(^2\)The supporting calculus is provided by Raiffa and Schlaifer, pp. 294-297.
\[ L = \begin{cases} k_0 (R-M) & \text{for } R \geq M \\ k_u (M-R) & \text{for } M > R \end{cases} \]  \hspace{1cm} (9)

where \( k_0 \) is the conditional loss per dollar of overstatement when \( R \) exceeds \( M \), and \( k_u \) is the conditional loss per dollar of understatement. A quadratic loss function is given by:

\[ L = k_t (R-M)^2 \]  \hspace{1cm} (10)

where \( k_t \) is a constant representing the loss per squared dollar of error. Finally, a step loss function is given by:

\[ L = \begin{cases} k_a & \text{if } R > M+E \\ k_b & \text{if } R < M-E \\ 0 & \text{otherwise} \end{cases} \]  \hspace{1cm} (11)

where \( k_a \) is a fixed dollar loss incurred if the estimate \( R \) exceeds the true amount \( M \) by more than a material amount \( E \), and \( k_b \) is a fixed dollar loss incurred if \( R \) is less than \( M-E \). Models incorporating each of these three forms of loss function are developed in this paper.

The author elected to use all three of these forms of loss function because the available evidence concerning the true form of accountant's and auditor's loss functions is both limited and inconclusive. In Figure 1 these three forms of loss function are graphically compared. It is clear from the graph that each of these loss functions differs radically from the others. In support of the selection of these particular forms, the author speculates that their substantial differences make them representative of the much wider
set of loss function forms which might actually exist among accountants and auditors. By varying the parameters of the loss function, at least one of these forms should provide a reasonably close approximation to many true loss functions. However, even where such an approximation is not satisfactory, it may still be possible to determine an optimal sample size for a non-standard loss function using a heuristic procedure similar to the technique described later in this paper for the step loss function.

Insert Figure 1 here

Discussions of loss functions which have appeared in the accounting literature, including Kinney [1975a, p. 119], Ward [1976, pp. 30-32], Scott [1975], and Felix and Grimald [1976], have all dealt with auditor's loss functions. Both Kinney and Felix and Grimald postulate that the auditor's loss consists of the expected costs of lawsuits, including both legal expenses and settlement costs, and of loss of the auditor's professional reputation.

To the accountant, or to the organization which employs the accountant, the nature of the loss function depends upon whether the estimate is being made for external or internal reporting purposes. In the latter case, loss would represent the cost of incorrect management decisions which could possibly be made as a result of the estimation error. In the case of external reporting, the accountant's loss would probably be similar in composition to the auditor's, in that it would consist of the expected cost of potential lawsuits together with the loss of credibility among those who use financial statement information. Note that in the case of both the accountant and auditor, the expected loss depends upon the probability that an error is alleged and/or revealed, or the probability that an incorrect decision is made, as well as upon the resulting consequences to the auditor, accountant, or organization.
Figure 1
Comparison of Linear, Quadratic, and Step Loss Functions
Scott [1975] has proposed a technique for deriving an auditor's loss function from an assumed investor decision model. His experiments show loss functions which are quadratic in form. This result could be a significant breakthrough if one could accept his approach and assumptions. However, the objections raised in discussion comments by Magee [1975] appear quite serious. Further, the issue of how such factors as legal liability, legal and insurance costs, damage to professional reputation, and possible loss of clients are factored into the derived auditor's loss function is not addressed by Scott. These questions must be dealt with before it could be concluded that the derived loss function is an appropriate tool. In this paper it is assumed that loss refers only to direct consequences to be incurred by the decision-maker (auditor, accountant, or organization), rather than to indirectly derived consequences as suggested by Scott.

Using a sample of 24 managers and partners in three national CPA firms, Ward [1976] obtained judgements about the relationship between the size of a financial statement error and the economic consequences to the auditor. Among the forms which he suggested to his subjects were the linear, exponential, and step loss functions. His subjects displayed very little agreement with respect to the appropriate functional form. However, Ward reports that over 60% of his subjects indicated that the impact of a material error levels off after a point, suggesting that there exists "some upper threshold which limits exposure to economic loss by the auditor as a result of a subsequently discovered error" [p. 31]. Such a concept is consistent with the step loss function, but not with either the linear or quadratic loss functions.

Further support for the assumption that the auditor's loss function is appropriately represented by a step function in provided by Ken Stringer, Partner in Charge of Accounting and Auditing Services for Deloitte, Haskins
Sells:

... my own feeling, and I believe that of many auditors is that once an auditor has passed that threshold of materiality, he is in trouble. Whether he is in linear trouble or exponential trouble may be of interest, but I am not sure that it is too important. [1976, pp. 47-48]

Another significant issue involves whether or not the accountant’s or auditor’s loss function is symmetric, that is, whether overstatements (of assets or income) have the same economic consequences as understatements. Ward’s 24 subjects were evenly divided on this issue. However, in discussion comments on the Ward paper, Kaplan [1976] rejects the idea of an asymmetric loss function, pointing out that auditors should be equally concerned with the interests of the seller and buyer of investment securities. However, it could be argued that the courts have shown greater concern for overstatements than understatements, and it is not clear whether a normative approach such as Kaplan suggests should overlook this tendency. A complicating factor is that whenever the cost of overstatement is assumed to exceed the cost of understatement, the optimal terminal point estimate recommended by the models will be less than the accountant’s or auditor’s maximum likelihood estimate of the quantity, and the accountant or auditor could be accused of dishonesty for reporting an amount less than he believes the correct value to be. Of course such reporting practices have long been justified on the basis of conservatism, and if the true loss functions of accountants and auditors are indeed asymmetric, then conservatism is clearly a rational form of behavior.

In conclusion the loss function issue is certainly the most complex problem faced in developing the models in this paper, and the available evidence is very inconclusive. Aside from the matter of determining the appropriate form of the loss function, there remains the problem of estimating the correct parameters of the loss function. Because of the uncertainty surrounding
these questions, the author considered it appropriate to develop one model for sample size determination and posterior point estimation which does not attempt to optimize costs and losses, and therefore does not require the specification of a loss function. This model, together with the three separate models incorporating linear, quadratic, and step losses, constitute the four models which are detailed in the subsequent section concerning sample size determination.

The Sampling Cost Function

As mentioned earlier it is assumed that the sampling cost function is linear and has fixed and variable components. Total sampling cost $T_s$ may then be expressed as:

$$T_s = K_s + k_s n$$

(12)

where $K_s$ is the fixed cost of preparing a sample plan and evaluating sample results, and $k_s$ is the cost of evaluating an average sample item, which involves investigation of documents, computation, follow-up on discovered errors in book values, follow-up on nonresponses to confirmations, etc. It is assumed that $k_s$ is constant across all sample strata.

The sampling cost function appears to be the least controversial aspect of the model proposed in this paper. Equation (12) has been consistently used in the accounting literature by all authors whose models incorporate a sampling cost function, including Moriarity [1973], Kinney [1975a, 1975b], Heimann and Chesley [1977], Kinney and Warren [1979], and Cushing, Searfoss and Randall [1979]. In fact both Kinney and Warren and Cushing, Searfoss and Randall were able to estimate variable sampling costs for actual audit applications based upon data provided by auditors.
Sample Size Determination

Combining the sampling cost and expected loss functions of the accountant or auditor, we obtain an expected total cost function as follows:

\[ E(TC) = T_s + E(L) \]  \hspace{1cm} (13)

where \( E(L) \) represents the product of the probability and loss functions.

The objective of the accountant or auditor is assumed to be the minimization of this expected total cost function. The treatment of \( T_s \) and \( E(L) \) as additive requires the assumption that the accountant or auditor is risk-neutral. The advantage of making this assumption, of course, is that it greatly simplifies the analysis. However, risk-neutrality is a rational assumption in this case whenever the magnitude of \( E(TC) \) is immaterial relative to the total wealth of the organization. This condition should hold in large corporations or public accounting firms.

The first step in the application of the model is to determine the sample size, \( n \), which minimizes \( E(TC) \). Since \( T_s \) clearly increases at a constant rate with respect to \( n \), and \( E(L) \) decreases at a decreasing rate as \( n \) increases, \( E(TC) \) is a U-shaped function with respect to \( n \). Conventional techniques of calculus may therefore be used to find the value of \( n \) which minimizes \( E(TC) \). In this section the required analysis is presented for the linear, quadratic, and step loss models, respectively. The section concludes with the presentation of a sample size determination model based upon a specified posterior interval width.

**Linear loss model.** In the linear loss case, for a given terminal point estimate \( R \) and probability function \( f(M) = f'(M) \) or \( f''(M) \) (prior or posterior), the expected terminal loss is given by:

\[ E(L^2) = k_s \int_{-\infty}^{R} (R - M)f(M) \, dM + k_u \int_{R}^{\infty} (M-R) \, f(M) \, dM \]  \hspace{1cm} (14)
It is first necessary to determine the optimal terminal point estimate $R^*$. As explained by Raiffa and Schlaifer [1961, p. 196], among others, when the loss function is of the form of equation (9), regardless of the form of the probability function, the optimal point estimate is given by the $k_r$th fractile of the probability distribution, where

$$k_r = \frac{k_u}{k_u + k_0}$$  \hspace{1cm} (15)

Therefore, since $f(M)$ is normal, we have

$$P(\tilde{M} < R^*) = P_N(\tilde{u} < u^*) = k_r$$  \hspace{1cm} (16)

where $\tilde{u}$ is the unit random variable corresponding to $\tilde{M}$, and $P_N(\tilde{u} < u^*)$ represents the portion of the cumulative unit normal distribution to the left of $u^*$. Given $k_r$, the value of $u^*$ may be obtained by table look-up, or in a computer program, by a simple numerical analysis routine. Once $u^*$ is determined, the optimal terminal point estimate $R^*$ is given by:

$$R^* = E(\tilde{M}) + u^*\sigma(\tilde{M})$$  \hspace{1cm} (17)

where under a prior distribution $E(\tilde{M}) = \bar{M}'$ and $\sigma(\tilde{M}) = \sigma'$. and under a posterior distribution $E(\tilde{M}) = \bar{M}''$ and $\sigma(\tilde{M}) = \sigma''$.

An expanded expression for expected total cost is:

$$E(TC) = K_S + k_s n + (k_0 + k_u) f_N(u^*) \sigma(\tilde{M})$$  \hspace{1cm} (18)

where the third term on the right represents $E(L^2)$ from equation (14) evaluated at $R = R^*$, and where $f_N(u^*)$ is the value of the unit normal distribution corresponding to $R^*$, i.e.

\textsuperscript{3}See, for example, Abramowitz and Stegun [1964, p. 932].
\[ f_N(u^*) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^*^2}{2}} \]

(19)

Next, an expression for \( \sigma(\tilde{M}) = N\sigma_n^* \) in terms of \( n \), based upon equations (7) and (5), must be substituted into equation (18); this yields:

\[
E(\text{TC}) = K_s + k_s n + (k_o + k_u) f_N(u^*) \frac{1}{\left[ \frac{1}{\sigma^2} + \frac{n}{(\Sigma h \bar{w}_h S_h)^2} \right]^k} \]

(20)

where \( V(\bar{Y}_{st}) = (\Sigma h \bar{w}_h S_h)^2 / n \) does not incorporate the finite population correction factor. Differentiating with respect to \( n \) and then solving for \( n \), we obtain

\[
n^0 = \left[ \frac{(\Sigma h \bar{w}_h S_h)}{2k_s} \left( k_u + k_o \right) N f_N(u^*) \right]^{2/3} \left( \frac{\Sigma h \bar{w}_h S_h}{\sigma^2} \right)^{2/3} \]

(21)

Since the finite population correction factor is not incorporated into equation (20), the correct value of \( n^0 \) will be slightly lower than given by (21). To find this correct value, substitute slightly smaller values of \( n \) into equation (20) until a value is found which minimizes that expression.\(^4\)

The optimal sample size \( n^* \) is either equal to \( n^0 \) or 0, depending on whether or not \( E(\text{TC}|n=n^0) \) is less than \( E(\text{TC}|n=0) \). If \( E(\text{TC}|n=n^0) \) is less than the terminal expected loss obtained by basing the point estimate on the prior distribution without sampling, then \( n^* = n^0 \); otherwise \( n^* = 0 \).\(^5\) If \( n^* = n^0 \), then the sample items may be allocated among strata using equation (4).

\(^4\)Incorporating the finite population correction factor makes the expression for \( n^0 \) extremely complex. Since \( E(\text{TC}) \) is known to be U-shaped for \( n>0 \), this heuristic procedure is effective and efficient for computer processing.

\(^5\)In some cases constraints may exist which require a nonzero sample size, as in the case of auditing standards with respect to receivables and inventories. In such cases it will always be true that \( n^* = n^0 \).
Quadratic loss model. Since both the probability function and the quadratic loss function given by equation (10) are symmetric, the optimal terminal point estimate \( R^* \) will always be equal to \( E(\tilde{M}) \), the mean of the prior or posterior distribution of the accountant or auditor. Given \( R^* = R = E(\tilde{M}) \), the expected total cost function reduces to:

\[
E(\text{TC}) = K_s + k_s n + k_t \sigma^2(\tilde{M})
\]

(22)

Again, in order to solve for \( n^0 \) we must substitute the complete expression for \( \sigma(\tilde{M}) = N \sigma'' \) into equation (22), which yields:

\[
E(\text{TC}) = K_s + k_s n + k_t N^2 \\
\left[ \frac{1}{\sigma^2} + \frac{1}{\frac{(\Sigma_h \hat{w}_h S_h)^2}{N} - \frac{\Sigma_h \hat{w}_h S_h}{N}} \right]
\]

(23)

Note that here \( V(\bar{Y}_{st}) \) does incorporate the finite population correction factor. Using calculus, we obtain a quadratic expression in \( n \), which is simplified by applying the familiar expression for the roots of a quadratic equation:

\[
n^0 = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}
\]

(24)

where:

\[
a = \frac{(\Sigma_h \hat{w}_h S_h^2)}{N^2 \sigma''} - \frac{2 \Sigma_h \hat{w}_h S_h^2}{N \sigma''^2} + 1
\]

(25)

\[^5\text{In theory an asymmetric quadratic loss function could be used; however, subsequent mathematical operations then become extremely cumbersome.}\]

\[^7\text{In the context of this model, the '+' or '-' in this equation should always be a '+'}.\]
\[
b = \frac{2(\Sigma_{h} \bar{W}_{h} S_{h})^{2}}{\sigma^{2}} = \frac{2(\Sigma_{h} \bar{W}_{h} S_{h})^{2} (\Sigma_{h} \bar{W}_{h} S_{h})^{2}}{N \sigma^{4}} \quad (26)
\]

\[
c = \frac{(\Sigma_{h} \bar{W}_{h} S_{h})^{4}}{\sigma^{4}} = \frac{k_{t} N^{2} (\Sigma_{h} \bar{W}_{h} S_{h})^{2}}{k_{s}} \quad (27)
\]

As in the linear loss case, a comparison of expected total cost at sample sizes of 0 and \(n^0\) is required in order to determine whether \(n^* = 0\) or \(n^0\).

Step loss model. In the step loss case, for given values of \(R\), \(E\), and \(f(M)\), expected terminal loss is given by:

\[
E(L^S) = k_{a} \int_{-\infty}^{R-E} f(M) dM + 0 \int_{R-E}^{R+E} f(M) dM + k_{b} \int_{R+E}^{\infty} f(M) dM \quad (28)
\]

This may be reduced to:

\[
E(L^S) = k_{a} P_{N}(\bar{M} \leq R-E) + k_{b} P_{N}(\bar{M} \geq R+E) \quad (29)
\]

The optimal terminal point estimate \(R^*\) will be less than \(E(\bar{M})\) if \(k_{a} > k_{b}\), equal to \(E(\bar{M})\) if \(k_{a} = k_{b}\), and greater than \(E(\bar{M})\) if \(k_{a} < k_{b}\). Now convert equation (29) into its equivalent in terms of the unit normal distribution:

\[
E(L^S) = k_{a} P_{N}(\bar{u} < u_{a}) + k_{b} P_{N}(\bar{u} > u_{b}) \quad (30)
\]

where

\[
u_{a} = \frac{R-E-E(\bar{M})}{\sigma(\bar{M})} \quad (31)
\]

\[
u_{b} = \frac{R+E-E(\bar{M})}{\sigma(\bar{M})} \quad (32)
\]
Since \( E(L^S) \) is clearly U-shaped with respect to \( R \), the value of \( R^* \) for \( k_a \neq k_b \) may be easily found by a computer program using a binary search technique to "zero-in" on the value of \( R \) which minimizes \( E(L^S) \).

To obtain the full expression for \( E(TC) \) in terms of \( n \), we must substitute the complete expression for \( U(\bar{m}) = No'' \) into (31) and (32), then in turn substitute these into (30), which is in turn substituted into the full expected total cost function. The resulting expression, although differentiable, does not enable a tractible equation for \( n^0 \) to be obtained, and so a heuristic procedure must be used to find \( n^0 \). Again, since \( E(TC) \) is U-shaped with respect to \( n \), a binary search procedure will accomplish this. Select an arbitrary value of \( n \), assume that \( \bar{Y}_{st} = m' \), solve for \( \sigma(\bar{m}) \), use the search procedure to find \( R^* \), and then determine the associated values of \( E(L^S) \) and finally \( E(TC) \). Repeat this procedure for systematically selected values of \( n \), zeroing in on the value \( n^0 \) which minimizes \( E(TC) \). As before, a check of this \( E(TC) \) value against \( E(TC|n=0) \) is necessary in order to determine whether \( n^* = 0 \) or \( n^0 \).

**Interval estimation model.** Many accountants may be reluctant to use the methods developed up to this point because of the difficulty and subjectivity involved in establishing loss parameters for the models. A procedure which might have more intuitive appeal is an interval estimation approach, which is simply a Bayesian extension of classical interval estimation techniques commonly used for variables estimation by accountants and

---

\(^6\)Because of its length, that expression is not shown here. However, the method described here for constructing it is completely analogous to the procedures illustrated earlier for constructing this function in the quadratic and linear loss cases. (See equations (20) and (23)).

\(^9\)Expected total cost is independent of the sample mean, so any arbitrary value for \( \bar{Y}_{st} \) may be assumed; the assumption given here is used simply for convenience.
auditors, and which does not consider costs or losses. However, instead of a classical two-sided confidence interval, the accountant substitutes a two-sided credible interval derived from his or her posterior probability distribution.

To apply this approach, the accountant must specify a desired level of confidence and a desired posterior interval width. In auditing, official pronouncements suggest that selection of a confidence level be based on the degree of risk of material error in the account, and that selection of the appropriate interval width be based on materiality considerations.\(^{16}\) For accounting applications in general, these guidelines may be appropriate as a starting point, and may be modified according to the situation and the judgement of the accountant.

Refer to the confidence level specified by the accountant as \( P \) and the desired total interval width as \( 2A \). Once these values are obtained a standard table of the unit normal distribution is consulted to find the value of \( u \) corresponding to \( P \) for a two-sided interval. Then the required per-unit posterior standard deviation is

\[
\sigma'' = \frac{A}{N/u} \quad (33)
\]

Next, substitute this expression and the expression for \( V(\bar{y}_{st}) \) from equation (5) into equation (7). For given values of \( \sigma'^2 \), \( \Sigma_h w_h s_h \) and \( \Sigma_h w_h s_h^2 \), \( n \) is the only remaining unknown. If the finite population correction factor is ignored, sample size is given by:

\[
\begin{align*}
\hat{n} & = \frac{u^2N(\Sigma_h w_h s_h)^2}{A^2} - \frac{(\Sigma_h w_h s_h)^2}{\sigma'^2} \\
& \quad (34)
\end{align*}
\]

\(^{16}\) American Institute of Certified Public Accountants. Statement on Auditing Standards No. 1, Section 320A, paragraphs .10-.13. This suggestion is directed at applications of classical statistics, in which the confidence level is referred to as the "reliability" and the interval width as "precision." Reasoning by analogy, the suggestion would seem to be appropriate for applications of Bayesian methods.
This is a useful intermediate result in computing the value of \( n \) which does incorporate the finite population correction factor:

\[
\begin{align*}
n &= \frac{(\sum W_h S_h)^2 N n^0}{(\sum W_h S_h)^2 N + (\sum W_h S_h)^2 n^0} \\
\end{align*}
\]  

(35)

As before, once this value of \( n \) is determined, it must be allocated among the individual strata using equation (4).

Once the sample has been taken and evaluated, a P\% posterior credible interval for the per unit population mean is given by \( m'' \pm u\sigma'' \), where \( \sigma'' \) and \( m'' \) are computed from equations (7) and (8), respectively. However, since a loss function is no longer incorporated into the model, there is no prescribed optimal point estimate \( R^* \). If the accountant feels that over- and under-estimates are equally undesirable, then he should select \( R = E(\hat{M}) \). Otherwise, he must use his judgement together with the information contained in the posterior credible interval to make this choice.

II. THE SENSITIVITY ANALYSES

The sensitivity tests reported in Part II were designed to answer the following questions in connection with the use by accountants or auditors of any of the optimization models described in Part I.

(1) If the accountant uses (a) sample sizes, or (b) terminal point estimates which differ from the optimal values, \( n^* \) and \( R^* \) respectively, prescribed by the model, how serious are the consequences of such deviations likely to be in terms of expected total cost?

(2) If the accountant specifies incorrect values of required model parameters, in particular (1) a prior mean \( m' \) differing from
the true population mean, (b) an incorrectly assessed prior standard deviation $\sigma'$, 11 (c) an incorrect sampling cost parameter $k_s$, (d) an incorrect loss parameter or parameters, $k_o$, $k_u$, $k_t$, $k_a$, or $k_b$, depending on the form of loss function, or (e) incorrect measures of population variability $\Sigma_h w_h s_h$ and $\Sigma_h w_h s_h^2 / N$, how serious are the consequences to the accountant likely to be in terms of expected total cost?

For convenience, the first set of questions is referred to as "policy issues", and the second as "parameter estimation issues." The purpose of these tests is to provide guidance to the accountant or auditor in estimating model parameters and interpreting model outputs.

A computer program was prepared which required the following input parameters, (1) population size, (2) fixed and variable sampling costs, (3) prior distribution standard deviation, (4) population variability measures, (5) population mean per unit, (6) code for linear, quadratic, or step loss, (7) loss function parameters, and (8) in the step loss case, the minimum amount of estimation error which would be considered material. For a given set of input parameters, the program solved for optimal sample size and the associated minimum expected total cost. Then, one at a time, the two policy variables were varied by specified percentages from the optimal values prescribed by the model, and the resulting increase in expected total cost, as a percentage of the minimum expected total cost, was printed out. Next, one by one, the five model parameters listed in research question (2) were each varied by specified percentages from their initially entered "correct" values, and

11An error in the assessed value of $\sigma'$ could arise from improper application of the assessment procedure, or from conversion of the assessed distribution into a more tractable form, such as by using the normal approximation technique proposed by Cushing and Romney (1980).
the model was solved for the policy which would have been prescribed using the "incorrect" parameter. In each case, the impact of using this "incorrect" policy, measured in terms of the resulting increase in expected total cost as a percentage of the previously determined "correct" minimum expected total cost, was computed and printed. Using a variety of input parameters (to be explained shortly), this program was run 32 times for each of the three loss function forms, or a total of 96 times. Then, for each of the research questions, the average and highest values of the percentage deviations from minimum expected total cost were determined and tabulated to prepare the tables shown later in Part II.

The tests of research questions 1b and 2a require the determination of expected terminal loss when the terminal point estimate differs from the optimal estimate $R^*$. Refer to the actual terminal point estimate as $Q$. In the step loss case, equation (30) is used for this purpose, with $u_a$ and $u_b$ determined by substituting $Q$ for $R$ in equations (31) and (32), respectively. However, in both the linear and quadratic loss cases, entirely different equations for expected terminal loss must be used. In the quadratic loss case, the appropriate equation is:\(^{12}\)

$$E(L^q | Q \neq R^*) = k \left[ (Q - E(R))^2 + \sigma^2(\bar{M}) \right] \tag{36}$$

In the linear loss case, the required formula is:\(^{13}\)

$$E(L^k | Q \neq R^*) = (k_o + k_u) \bar{f}_N(u_q) \sigma(\bar{M}) + \left[ k_o \bar{P}_N(\tilde{u}_q \leq u_q) - k_u \bar{P}_N(\tilde{u}_q > u_q) \right] u_q \sigma(\bar{M}) \tag{37}$$

---

\(^{12}\)Adapted from Raiffa and Schlaifer, 1961, p. 188, equation 6-26.

\(^{13}\)Adapted from Schlaifer, 1959, p. 301.
where:

\[ u = \frac{Q - E(\bar{y})}{\sigma(\bar{y})} \]  

(38)

In those instances where parameter errors induced a recommend sample size of less than 30 items, the computer program automatically set the sample size at 30. The assumption underlying this policy is that most accountants or auditors taking a stratified sample would want to use a sample of at least this many items in order to minimize the possibility of substantial error in estimating \( V(\bar{y}_{st}) \). Furthermore, values of \( n=0 \) were not allowed because it was assumed that the accountant or auditor would always wish to base his estimate at least partially on sample data.

**Selection of Test Parameters**

In order for the results of the sensitivity analyses to be meaningful, it is necessary to use a wide range of parameter values which are representative of those encountered in accounting practice. Toward this end, empirical data from three separate sources were consulted. These were (1) the Neter and Loebbecke [1975] study, (2) the Romney [1977] study of auditor’s prior distributions, and (3) the field tests reported by Cushing, Searfoss and Randall [1979]. In the following paragraphs, the choice of parameter values for the test cases is explained.

**Population size and mean.** The population size \( N \) and the population mean \( \mu \) generally impact only the magnitude of the expected total cost numbers, and have little or no impact upon the percentage deviations from these numbers which are tabulated here. Accordingly, only a single value of \( N \) was used in all test cases, that being 5000. This value is approximately the average of the four Neter and Loebbecke populations and twelve populations examined in the Cushing, Searfoss and Randall field tests.
Values of \( \mu \) equal to $1000 and $10000 were each used in one-half of the test cases. These values represent average means for small and large dollar value populations, respectively, among the sixteen populations referred to in the preceding paragraph.

Generally, values of \( m' \) and \( \overline{y}_{st} \), the prior and sample per-unit means, were not required by the models because the preposterior expected total cost in entirely independent of these values. The one exception to this is the investigation of deviations of \( m' \) from \( \mu \). For these tests it was assumed that \( \overline{y}_{st} = \mu \) since indeed \( \mu \) does represent the expected value of \( \overline{y}_{st} \).

Variable sampling cost. Cushing, Seafoss and Randall obtained estimates of variable sampling cost per unit for twelve estimation sampling applications. These values averaged about $12.50 and varied from $5.75 to $25.10, depending primarily upon the expected number of errors requiring follow-up work. However, these values reflect sampling costs incurred by auditors, and may not be representative of the sampling costs incurred for internal accounting applications. The latter are likely to be smaller because (1) they will reflect actual wages and salaries rather than audit billing rates, and (2) internal accountants assigned to prepare an estimate may be expected to do less follow-up or audit work than auditors performing an attest function. Accordingly, for one-half of the test cases a value of \( k_s \) equal to $12 was used, while a value of $6 was used in the remaining cases.

Population variability measures. Among the four Neter and Loebbecke populations, the value of \( \sum W_h S_h \) averaged about 3\% of the population mean. Among the twelve populations included in the Cushing, Seafoss and Randall study, the value of \( \sum W_h S_h \) ranged between 4\% and 13\% of the population mean. Accordingly, for half of the sensitivity tests a value of \( \sum W_h S_h \) equal to 4\% of \( \mu \) was used, and for the other half a value equal to 12\% of \( \mu \) was used.
Also, based upon data relationships found to exist in several of these populations, the value of the finite population correction factor was set at .001
\[ x \left( \sum_{h} W_{h} S_{h} \right)^2 \] for all test cases.

**Prior standard deviation.** Among the subject auditors for whom subjective distributions were assessed by Romney, the prior distribution standard deviations varied between 2% and 12% of the population mean, and averaged about 3%. Accordingly, in one-half of the test cases the prior distribution standard deviation was set at 3% of \( \mu \). In the other half of the cases, the prior standard deviation was set at 1.5% of \( \mu \). The possibility of running a set of cases for which the ratio \( \sigma' / \mu \) was 6% or higher was rejected because the prior distribution was so diffuse in these cases that it had little effect on the results. This suggests that incorporation of the prior distribution into the analysis is not worthwhile unless the accountant's confidence in the prior information is relatively high.

**Equivalent prior sample sizes.** The reduction in optimal sample size attributable to the incorporation of a prior probability distribution into the analysis is referred to as the equivalent prior sample size. This value, designated \( n' \), is approximated by

\[ n' = \left( \sum_{h} W_{h} S_{h} \right)^2 / \sigma'^2 \quad (39) \]

This approximation does not take into account the finite population correction factor, and as a result the true value of \( n' \) will always be slightly larger than the value given by this approximation.

The values chosen for \( \sum_{h} W_{h} S_{h} \) and \( \sigma' \), as described above, determine the values of \( n' \) for the test cases. The resulting values of \( n' \) are 1.78, 7.11, 16, and 64, each of which is applicable to one-fourth of test cases. To the extent that the values of \( \sum_{h} W_{h} S_{h} \) and \( \sigma' \) chosen here are representative of actual
values, these equivalent prior sample sizes are about what might be expected in practice. As mentioned earlier, one of the advantages of the Bayesian approach is the reduction in sample size permitted when prior information is formally incorporated into the analysis. These numbers suggest that the magnitude of this reduction may often be moderate or very small when the stratified mean-per-unit estimator is used.

Loss function parameters. Unfortunately, no evidence exists concerning appropriate values of loss function parameters for any of the three functional forms of loss. However, since all of the other model parameter values used have empirical support, it was decided to select loss parameters which resulted in sample sizes within the range of 50 to 500, which represents the range of sample sizes generally encountered in accounting and auditing practice. To the extent that the loss function forms and values of other parameters tested are representative of true conditions, the loss function parameters obtained in this way will be of the same order of magnitude as "true" loss parameters implicit in the use of such sample sizes. For each of the sixteen possible combinations of the other four varied parameters specified above, two different loss function parameters were obtained in this manner for each functional form of loss. This resulted in the total of 32 test cases for each of the three forms of loss function, or 96 cases in all.

In all of the test cases a symmetric loss function was used. The author did experiment with numerous asymmetric loss functions in the linear loss case, and found that this factor had virtually no impact upon the conclusions reached from the sensitivity tests.

The step loss function is not complete without specification of $E$, the threshold of material error. Among auditors, the most widely mentioned rule of thumb is that materiality is roughly 5% of net income. In most cases, the
balances of such asset accounts as inventory, accounts receivable, and fixed assets are greater than net income, often substantially, and so a materiality percentage of less than 5% of such balances might be appropriate. Leslie [1977, p.13] cites Newman's reference to precision values of 3 to 5% of an account balance; he then disputes the adequacy of such percentages, suggesting instead 1/2%. When a materiality percentage of 2% or higher was used in the step loss model, enormous and seemingly unrealistic values of the loss parameter were required in order to obtain optimal sample sizes larger than 50. Accordingly, a value of E equal to 1% of the population dollar total was used in all of the step loss test cases.

Results: Policy Issues

The average and highest values of the percentage opportunity loss for the 32 test cases run for each of the three loss functions are tabulated in Table 1 for the two policy issues. These results display a dramatically consistent pattern over the wide range of parameters and loss function forms tested. First, they show that the expected total cost function is generally flat in the neighborhood of the optimal sample size. In the linear loss cases, deviations of n ranging from -50% to +90% of n* never resulted in an opportunity loss higher than 11% of the minimum expected total cost. The quadratic and step loss functions showed somewhat greater sensitivity, but still indicate the existence of a sizable range of values of n which are nearly optimal.

---

Insert Table 1 here

---

Upon reflection, this lack of sensitivity of expected total cost to sample size is not so surprising. Recall that expected total cost is the sum of sampling cost and expected loss. Sampling cost is an increasing linear function of sample size, while expected loss is a downward sloping convex function,
<table>
<thead>
<tr>
<th></th>
<th>Sample Size</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%Deviation from Optimal Value</td>
<td>%Opportunity Loss Average</td>
<td>%Opportunity Loss Highest</td>
<td>%Deviation from Optimal Value</td>
<td>%Opportunity Loss Average</td>
<td>%Opportunity Loss Highest</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>LINEAR</strong></td>
<td>-50%</td>
<td>7.60</td>
<td>9.97</td>
<td>± 1%</td>
<td>122</td>
<td>341</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-10%</td>
<td>0.20</td>
<td>0.25</td>
<td>± 5%</td>
<td>803</td>
<td>1886</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+10%</td>
<td>0.17</td>
<td>0.22</td>
<td>±10%</td>
<td>1656</td>
<td>3818</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+50%</td>
<td>3.28</td>
<td>4.14</td>
<td>±20%</td>
<td>3363</td>
<td>7682</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+90%</td>
<td>8.39</td>
<td>10.85</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>LOSS</strong></td>
<td>-50%</td>
<td>16.73</td>
<td>23.63</td>
<td>± 1%</td>
<td>304</td>
<td>1095</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-10%</td>
<td>0.40</td>
<td>0.54</td>
<td>± 5%</td>
<td>7598</td>
<td>27385</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+10%</td>
<td>0.34</td>
<td>0.45</td>
<td>±10%</td>
<td>30394</td>
<td>109542</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+50%</td>
<td>6.30</td>
<td>8.08</td>
<td>±20%</td>
<td>121575</td>
<td>438168</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+90%</td>
<td>16.21</td>
<td>20.49</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CASES</strong></td>
<td>-50%</td>
<td>66.00</td>
<td>166.59</td>
<td>± 1%</td>
<td>1125</td>
<td>4892</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-10%</td>
<td>1.36</td>
<td>2.62</td>
<td>± 5%</td>
<td>2266</td>
<td>9791</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+10%</td>
<td>1.01</td>
<td>2.01</td>
<td>±10%</td>
<td>2262</td>
<td>9791</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-50%</td>
<td>15.56</td>
<td>26.70</td>
<td>±20%</td>
<td>2262</td>
<td>9791</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>+50%</td>
<td>54.51</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
the slope of which gradually decreases in absolute value. Under these conditions, expected loss is nearly linear over wide ranges of sample size values, and therefore expected total cost will be relatively flat over a wide range of values in the neighborhood of optimal sample size. Furthermore, note that these conditions should exist regardless of the choice of statistical estimator, because the general relationship between sample size and variance is the same for all classical statistical estimators. In other words, the lack of sensitivity of expected total cost to deviations from optimal sample size may be expected to exist whenever sampling cost is linear and a classical statistical estimator is used.

This conclusion is quite significant, because it suggests that strenuous efforts to determine exact optimal sample sizes may not be warranted. In particular, techniques which sacrifice analytical accuracy to gain mathematical tractability in the determination of sample sizes should generally be preferred to techniques which do the opposite. For example, this conclusion reflects favorably on the assumption made earlier that accountant’s or auditor’s prior distributions may be reasonably approximated by normal distributions. Furthermore, it suggests that if loss function forms and parameters are difficult to determine, an interval estimation model such as the one developed at the conclusion of Part I may be superior to a more complete model which computes optimal sample sizes.

The results with respect to deviations from the optimal terminal point estimate \( R^* \) tell a different story. They indicate that expected total cost is extremely sensitive to the use of \( R \) values deviating from \( R^* \) by 1% or more. For example, in the linear loss cases a deviation of 1% from \( R^* \) resulted on average in a 122% increase in the expected total cost. The quadratic and step loss cases show even greater sensitivity. This indicates that
the accountant must be very cautious about selecting a terminal point estimate different from $R^*$. One situation for which this result has an interesting implication is where the auditor is using a statistical test to confirm the accuracy of a recorded book value. Applying either classical statistics or dollar unit sampling, the auditor should be satisfied with the book value whenever it falls anywhere within his confidence interval. In such cases no adjustment to the book value, other than for corrections of errors found in sample items, would be considered. However, the decision-theoretic model implies that whenever $R^*$ deviates from book value, even though book value may lie within the posterior credible interval, an adjustment of book value to an amount close to or equalling $R^*$ should be made. If for some reason such an adjustment is not feasible, or the need for it is disputed, then the best remaining option for the auditor may be to expand the sample to obtain more precise information which would either confirm the accuracy of the book value or justify an adjustment.

More generally, this result suggests that any technique which introduces a bias into the terminal point estimate may significantly increase expected total cost, even though the bias may be small. Accountants and auditors have generally ignored the bias present in the ratio and regression estimators under the assumption that its negligible size indicated a negligible effect. Perhaps this assumption is not warranted. In this light the lack of bias of the difference and mean-per-unit estimators may be viewed as a more significant advantage than previously supposed.

Finally, note that if there is a systematic bias in accountant's or auditor's prior distribution means, this will also introduce bias into the terminal point estimate. Empirical research will be necessary to determine
whether accountant's or auditor's priors for account balances do display such a systematic bias, or whether a normal approximation technique such as that proposed by Cushing and Romney might tend to introduce such a bias. If such a bias is found to exist to any significant degree, it could call into question the entire concept of incorporating a prior distribution into the accountant's or auditor's decision analysis. In other words, bias inherent in a prior distribution may be such a significant disadvantage as to more than offset the advantage of sample size reductions attributable to the prior. This analysis suggests that the issue of whether accountant's or auditor's priors are systematically biased is a significant research question.

**Results: Parameter Estimation Issues**

In light of the consistency and strength of the results with respect to the policy issues, some predictions may be made regarding the results of the parameter estimation tests. Specifically, expected total cost should not be very sensitive to estimation errors in those parameters which only affect the determination of sample size and do not affect the terminal point estimate; these include the variable cost of sampling and the loss function parameters. However, expected total cost may be sensitive to errors in those parameters which do impact upon the terminal point estimate, including the prior mean, the prior variance, and the population variance estimates.

Table 2 contains average and highest values of the percentage opportunity loss over 32 cases under each loss function with respect to the five parameter estimation issues. Once again the results are quite consistent. By far the most critical parameter in terms of estimation accuracy is the prior distribution mean. An error of 10% in estimating the prior mean could result in a percentage opportunity loss of from 49% to over 600%, depending upon the specific parameter values and form of loss function. Therefore, the
accountant or auditor who utilizes a Bayesian approach to variables estimation should devote careful attention to minimizing the bias of the prior mean relative to the true population mean.

---

Insert Table 2 here

Generally the results show that errors in estimating the variable sampling cost and the loss parameters are not very serious unless a substantial underestimate of 50% to 90% below the true value is made. Accountants should certainly be able to estimate sampling cost per unit to within a degree of accuracy much tighter than this. However, accountants or auditors may not be capable of consistently estimating the loss parameter this accurately. The results do show that for all three of these parameters, overestimation by up to 90% has virtually an inconsequential effect. The obvious conclusion is that the accountant or auditor should be quite conservative in estimating these parameters in order to minimize the possibility that he has underestimated them. Once again the results suggest that the conservatism frequently displayed by accountants and auditors has a rational economic basis.

Due to the design of the tests, the results with respect to errors in the prior and population variance are inconclusive. Errors in either of these parameters will contribute to a bias in the expected terminal point estimate only if there is error present in the prior mean. However, in these tests each variable was tested independently of the others; for example, the tests of effects of errors in the prior variance were run assuming that the prior mean was correct. Therefore, the results in Table 2 only reflect the effects of errors in prior and population variance estimates attributable to the incorrect sample sizes induced by these errors. To measure the effects on expected total cost when these errors occur jointly with an error in the prior mean would require a different test.
### Table 2
Parameter Error Sensitivity Results

<table>
<thead>
<tr>
<th>Deviation of Parameter from &quot;True&quot; Value</th>
<th>Prior Mean Average</th>
<th>Highest</th>
<th>Prior Variance Average</th>
<th>Highest</th>
<th>Sampling Cost Average</th>
<th>Highest</th>
<th>Loss Parameter Average</th>
<th>Highest</th>
<th>Population Variance Average</th>
<th>Highest</th>
</tr>
</thead>
<tbody>
<tr>
<td>LINEAR</td>
<td>-90%</td>
<td>716.0</td>
<td>2633.6</td>
<td>45.50</td>
<td>118.17</td>
<td>70.78</td>
<td>153.07</td>
<td>38.46</td>
<td>50.98</td>
<td>11.67</td>
</tr>
<tr>
<td></td>
<td>-50%</td>
<td>376.5</td>
<td>1436.2</td>
<td>4.46</td>
<td>27.32</td>
<td>6.52</td>
<td>9.73</td>
<td>5.42</td>
<td>7.50</td>
<td>1.76</td>
</tr>
<tr>
<td></td>
<td>-10%</td>
<td>49.3</td>
<td>238.7</td>
<td>0.02</td>
<td>0.11</td>
<td>0.16</td>
<td>0.27</td>
<td>0.14</td>
<td>0.23</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>+10%</td>
<td>49.3</td>
<td>238.7</td>
<td>0.01</td>
<td>0.10</td>
<td>0.12</td>
<td>0.18</td>
<td>0.13</td>
<td>0.23</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>+50%</td>
<td>376.5</td>
<td>1436.2</td>
<td>0.10</td>
<td>0.85</td>
<td>1.96</td>
<td>2.84</td>
<td>2.28</td>
<td>4.77</td>
<td>1.58</td>
</tr>
<tr>
<td></td>
<td>+90%</td>
<td>716.0</td>
<td>2633.6</td>
<td>0.16</td>
<td>1.35</td>
<td>6.69</td>
<td>6.51</td>
<td>5.60</td>
<td>9.12</td>
<td>3.94</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Deviation of Parameter from &quot;True&quot; Value</th>
<th>Prior Mean Average</th>
<th>Highest</th>
<th>Prior Variance Average</th>
<th>Highest</th>
<th>Sampling Cost Average</th>
<th>Highest</th>
<th>Loss Parameter Average</th>
<th>Highest</th>
<th>Population Variance Average</th>
<th>Highest</th>
</tr>
</thead>
<tbody>
<tr>
<td>QUADRATIC</td>
<td>-90%</td>
<td>9397.0</td>
<td>50709.3</td>
<td>230.71</td>
<td>686.97</td>
<td>67.02</td>
<td>81.55</td>
<td>58.71</td>
<td>76.61</td>
<td>138.19</td>
</tr>
<tr>
<td></td>
<td>-50%</td>
<td>2900.3</td>
<td>15631.1</td>
<td>10.08</td>
<td>37.18</td>
<td>5.46</td>
<td>6.73</td>
<td>5.43</td>
<td>6.63</td>
<td>15.61</td>
</tr>
<tr>
<td></td>
<td>-10%</td>
<td>116.0</td>
<td>626.0</td>
<td>0.03</td>
<td>0.36</td>
<td>0.12</td>
<td>0.15</td>
<td>0.12</td>
<td>0.15</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>+10%</td>
<td>116.0</td>
<td>626.0</td>
<td>0.02</td>
<td>0.11</td>
<td>0.10</td>
<td>0.13</td>
<td>0.10</td>
<td>0.13</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>+50%</td>
<td>2900.3</td>
<td>15631.1</td>
<td>0.19</td>
<td>1.36</td>
<td>1.55</td>
<td>2.31</td>
<td>1.35</td>
<td>2.27</td>
<td>5.41</td>
</tr>
<tr>
<td></td>
<td>+90%</td>
<td>9397.0</td>
<td>50709.3</td>
<td>0.30</td>
<td>2.13</td>
<td>4.70</td>
<td>5.82</td>
<td>4.63</td>
<td>5.70</td>
<td>13.35</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Deviation of Parameter from &quot;True&quot; Value</th>
<th>Prior Mean Average</th>
<th>Highest</th>
<th>Prior Variance Average</th>
<th>Highest</th>
<th>Sampling Cost Average</th>
<th>Highest</th>
<th>Loss Parameter Average</th>
<th>Highest</th>
<th>Population Variance Average</th>
<th>Highest</th>
</tr>
</thead>
<tbody>
<tr>
<td>IR2D</td>
<td>-90%</td>
<td>1750.7</td>
<td>9791.0</td>
<td>374.05</td>
<td>1437.43</td>
<td>16.13</td>
<td>45.11</td>
<td>39.48</td>
<td>157.19</td>
<td>374.03</td>
</tr>
<tr>
<td></td>
<td>-50%</td>
<td>1298.8</td>
<td>9781.3</td>
<td>4.76</td>
<td>29.63</td>
<td>2.76</td>
<td>5.54</td>
<td>3.22</td>
<td>12.94</td>
<td>150.11</td>
</tr>
<tr>
<td></td>
<td>-10%</td>
<td>103.4</td>
<td>233.3</td>
<td>0.02</td>
<td>0.19</td>
<td>0.08</td>
<td>0.26</td>
<td>0.09</td>
<td>0.28</td>
<td>2.23</td>
</tr>
<tr>
<td></td>
<td>+10%</td>
<td>103.4</td>
<td>233.3</td>
<td>0.01</td>
<td>0.08</td>
<td>0.08</td>
<td>0.21</td>
<td>0.07</td>
<td>0.21</td>
<td>1.49</td>
</tr>
<tr>
<td></td>
<td>+50%</td>
<td>1298.8</td>
<td>9781.3</td>
<td>0.10</td>
<td>0.77</td>
<td>1.57</td>
<td>4.29</td>
<td>1.09</td>
<td>3.34</td>
<td>19.30</td>
</tr>
<tr>
<td></td>
<td>+90%</td>
<td>1750.7</td>
<td>9791.0</td>
<td>0.16</td>
<td>1.20</td>
<td>4.35</td>
<td>11.03</td>
<td>2.42</td>
<td>7.43</td>
<td>40.71</td>
</tr>
</tbody>
</table>

***Tabled values represent percentage increases in expected total cost induced by incorrect measurement of the parameter value.***
Taking this into account, the results in Table 2 show that errors in estimating the prior variance, $\sigma^*$, have virtually no effect on expected total cost unless $\sigma^*$ is underestimated by 50% or more. Such gross underestimates are probably unlikely. However, if the prior mean is also in error, then any underestimate of the prior variance will increase the weight given to the prior mean in the calculation of the posterior mean, thereby contributing to additional error in the terminal point estimate and leading to potentially drastic increases in expected total cost. All of this suggests a conservative strategy of deliberately overestimating the prior variance. However, as indicated by equation (39) and the related discussion, such a strategy would reduce the equivalent prior sample sizes associated with prior information, thereby reducing or even eliminating the advantage of using the prior distribution in the first place.

With respect to the population variability measure $\sum \lambda \beta \delta$, the results indicate that estimation errors of $\pm 10\%$ should generally have only a minor effect on expected total cost. However, estimation errors in the range of $\pm 50\%$ or more may have serious consequences, particularly if the loss function is a step function. Furthermore, underestimation errors appear to be significantly more serious than overestimation errors. However, if the prior mean is also in error, then overestimation of the population variability will increase the weight given to the prior mean in the calculation of the posterior mean, thereby contributing to additional error in the terminal point estimate and increased opportunity loss. By the same token, underestimation of the population variability in the presence of an error in the prior mean would mitigate the effect of such an error on opportunity loss. All that may be concluded from this is that errors of overestimation or underestimation of population variability could have serious effects. To reach a more definite
conclusion would require not only an extension of the tests reported here, but also an awareness which we do not now possess of the magnitude of these errors likely to be encountered in practice.

In summary, the parameter for which estimation accuracy is the most important is the prior distribution mean, while estimation accuracy is least important for variable sampling cost. Estimation accuracy is also not critical for the loss function parameters as long as they are not grossly underestimated. Finally, our results are inconclusive with respect to both the prior and population variance estimates.

A Digression on Reliability and Risk

The Bayesian approach provides one method of taking into account prior information in the design of sampling plans. Another approach, suggested by the AICPA, [SAS 1, Section 32OB.35], is to specify a desired overall level of statistical reliability, and then to reduce this by a factor reflecting the prior information available concerning system reliability, analytical review, etc., according to the following formula:

\[ S = 1 - \frac{(1-R)}{(1-C)} \]  \hspace{1cm} (40)

where \( R \) is the desired overall reliability level, \( C \) is the reliance assigned to the internal control system and other factors about which prior information is available, and \( S \) is the reliability level to be used in designing the statistical test.\(^1\)

The major problem in applying equation (40) in practice has been the estimation of \( C \), the reliance assigned to prior information. Interestingly,

\(^1\)The logic of this formula is explained more fully by Roberts [1978, pp. 10-13], and by Ernst & Ernst [1976, pp. 66-76].
if certain assumptions are made, an implied value of \( C \) may be computed from the applications of the models used in this paper. The normal curve \( u \)-factor corresponding to the overall reliability \( R \) is given by:

\[
\frac{u_R}{R} = \frac{A/N}{\sigma''} \tag{41}
\]

where \( A/N \) is targeted statistical precision expressed on a per-unit basis, and \( \sigma'' \), of course, incorporates all sources of information concerning \( u \) (prior information plus statistical test results). The normal curve \( u \)-factor corresponding to the statistical test reliability \( S \) is in turn given by:

\[
\frac{u_S}{S} = \frac{A/N}{(V(\overline{y}_{st}))^{\frac{1}{2}}} \tag{42}
\]

Given \( u_R \) and \( u_S \), values of \( R \) and \( S \) may be obtained by table look-up or, within a computer program, by numerical analysis. Then the implied value of \( C \) corresponding to those values of \( R \) and \( S \) may be obtained from the following expression:

\[
C = 1 - \frac{(1-R)}{(1-S)} \tag{43}
\]

which is, of course, derived by algebraic manipulation of equation (40).

Since our test case parameters were chosen to be as realistic as possible in light of available evidence, the implied values of \( C \) for the test cases should also be a realistic reflection of the amount of weight which should be assigned to reliance on prior information by accountants or auditors. To compute the implied \( C \) value for each case requires that we have values for \( A, N, \sigma'' \), and \( V(\overline{y}_{st}) \). All of these values are available except for \( A \), so values of \( A \) must be assumed for each test case. To do this, the 32 test cases for each loss function are evenly divided into "high reliance"
cases, in which \( \sigma' \), the assumed prior distribution standard deviation, is a relatively tight 1.5% of the population mean, and "moderate reliance" cases in which \( \sigma' \) is equal to 3% of the population mean. No "low reliance" cases were included in the tests because prior distributions for which \( \sigma' \) is more than 3% of the population mean are so diffuse that they allow virtually no reduction in sample size. Next, "high reliance" on prior information is equated with a statistical reliability of 50%,\(^{15}\) which in turn implies that statistical precision is equal to materiality; in terms of our model this enables us to set \( A = E \). Furthermore, "low reliance" on prior information would equate to statistical reliability of 95% or more,\(^{16}\) or roughly \( 2A = E \). Splitting the difference, it seems reasonable to assume that "moderate reliance" would correspond approximately to \( 1.5A = E \), or \( A = 2/3 \ E \).

Given the \( A \) values derived from these assumptions, implied \( C \) values for each test case have been computed, and the results are summarized in Table 3. In light of conventional auditing practice, the results shown in Table 3 are startling. They suggest that for those cases where the auditor's confidence in his prior information may be characterized as moderate, the reliance factor for internal control and other prior information should generally not exceed 5%. Even for the high confidence cases, the results indicate that the reliance factor should generally not exceed 30%. These reliance factors are substantially lower than those suggested for use in audit practice. For example, Ernst and Ernst \([1976, \text{p. 69}]\) suggest that:

\(^{15}\)This is standard practice in auditing; for example see Elliott and Rogers, \([1972]\).

\(^{16}\)Ibid.
Generally, you should value internal control effectiveness in the range from 0% to 90%. A value higher than 90%, particularly in the range from 95% to 100%, would not be encountered too frequently, and could preclude the need for a statistical test.

The AICPA provides no suggestions concerning specific values of C, but does provide a table in which illustrative values ranging from 30% to 90% are shown [SAS 1, Section 320B.35] thereby implying that such values might be reasonable.

As indicated above, a primary factor influencing these results is the relative magnitude of $\sigma'$, the prior standard deviation. The values of $\sigma'$ used in the test cases were based upon prior distributions assessed by Romney [1977] from auditors who were given hypothetical cases. The lack of reality of these cases could have caused Romney's subjects to express less confidence in their estimates than they would have in real cases. Furthermore, auditors (and accountants too) may simply be more willing and able to express their prior information in terms of the parameter C than in terms of $\sigma'$.

On the other hand, these results should serve as a reminder that the information content of a sample item (in terms of variance reduction) under a stratified sampling plan is relatively high. Therefore, any interpretation of prior information which significantly reduces the size of a stratified sample (such as by assigning a high reliance factor to the internal control system) implies that the prior information has enormous information content. In other words, accountants or auditors who use equation (40) in sample planning should recognize that C is a very critical parameter which should be estimated with care.

---

17In fact, calibration tests performed by Romney [Chapter VI] on his subjects indicated that they were generally "underconfident".
Table 3

Implied C Values for Test Cases

<table>
<thead>
<tr>
<th></th>
<th>High Confidence Cases *</th>
<th>Moderate Confidence Cases*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lowest C =</td>
<td>0.2060</td>
<td>0.0240</td>
</tr>
<tr>
<td>Average C =</td>
<td>0.2326</td>
<td>0.0319</td>
</tr>
<tr>
<td>Highest C =</td>
<td>0.2852</td>
<td>0.0423</td>
</tr>
<tr>
<td>Lowest C =</td>
<td>0.2044</td>
<td>0.0256</td>
</tr>
<tr>
<td>Average C =</td>
<td>0.2354</td>
<td>0.0328</td>
</tr>
<tr>
<td>Highest C =</td>
<td>0.2807</td>
<td>0.0418</td>
</tr>
<tr>
<td>Lowest C =</td>
<td>0.2125</td>
<td>0.0280</td>
</tr>
<tr>
<td>Average C =</td>
<td>0.2252</td>
<td>0.0309</td>
</tr>
<tr>
<td>Highest C =</td>
<td>0.2645</td>
<td>0.0385</td>
</tr>
</tbody>
</table>

*For high confidence cases, $\sigma'/\mu = .015$ and $A=E$, while for moderate confidence cases, $\sigma'/\mu = .03$ and $A=2/3\ E$. 
The accountant or auditor who wishes to use equation (40) in lieu of a Bayesian approach could test the reasonableness of any particular C value by constructing the normal prior probability distribution implied by that C value and his estimate $m'$ of the population mean. This would require values of $A$ (precision), $N$, and desired overall reliability $R$. The procedure is as follows:

1. Find the required posterior standard deviation $\sigma''$ by substitution of $A$, $N$, and $u_R$ into equation (41).
2. For the value of $C$ to be tested, find the corresponding value of $S$ using equation (40).
3. Find the required sample variance $V(\bar{Y}_{st})$ by substitution of $A$, $N$, and $u_S$ into equation (42).
4. By substituting $\sigma''$ and $V(\bar{Y}_{st})$ into equation (7), the value of $\sigma'$ implied by $C$ may be determined.
5. Using $m'$ and $\sigma'$, prior probability statements concerning the population mean $\mu$ may be constructed, or a graph may be prepared.

Similar statements or a graph may be constructed for the population total $M = N\mu$ using $E(\bar{M}) = Nm'$ and $\sigma(\bar{M}) = N\sigma'$.

Taking into account the possibility that his true feelings about the value of $\mu$ may not be properly reflected by a normal distribution, the accountant or auditor may then evaluate the reasonableness of the prior probability distribution implied by the given C value. If the probability statements derived from the prior distribution seem reasonable in light of the available prior information, then the chosen C value may be appropriate. However, if those probability statements seem to reflect greater confidence than is warranted by the prior information, then a smaller value of C should be used.
III. SUMMARY

In accounting and auditing practice, the problem of estimating the value of an unknown variable, such as an account balance, arises frequently. Often, the accountant or auditor possesses a substantial amount of prior information concerning the quantity to be estimated. This information may be summarized in the form of a prior probability distribution over the quantity. Assuming use of the stratified mean-per-unit estimator, this paper develops four models, one of which may be chosen for sample size determination, depending upon the user's objectives. The paper shows how the prior and sample information may be combined to obtain posterior point and/or interval estimates.

The primary economic advantages of the models are that (1) the quantification of prior information enables the required sample size to be reduced, and (2) the sample sizes recommended by the optimization models minimize the sum of sampling cost and expected loss of incorrect estimates. The primary disadvantages are (1) the time required to obtain estimates of required model parameters, and (2) the highly subjective nature of some of these parameters, especially those relating to the loss function.

Using parameter values selected to be as realistic as possible, a number of test cases were constructed and used to perform an extensive sensitivity analysis of the models. The results of these tests show that expected total cost is relatively insensitive to deviations from optimal sample size, as well as to errors in the estimation of the variable cost of sampling and the loss function parameters. However, expected total cost is relatively sensitive to errors in assessing the prior distribution mean, and is extremely sensitive to deviations from the optimal terminal point estimate prescribed by the model. These results suggest that the development of more refined
models for optimal sample size determination may not be worthwhile, but that
greater attention should be given to obtaining terminal point estimates
which are as accurate as possible.

An alternative method of taking prior information into account in
sample planning, suggested by the AICPA, is to assign a reliance factor to
it, thereby reducing the reliance factor required for the statistical tests.
The data available from the sensitivity tests permitted an empirical compari-
son to be made of the AICPA method, which has come into widespread use in
auditing, and the formal Bayesian method. While the outcome is not conclu-
sive in light of the assumptions necessary to construct this comparison,
the results suggest that the reliance factors commonly assigned to prior
information in current auditing practice may be far too high.

For those who advocate the implementation of a formal Bayesian approach
to variables estimation in accounting and auditing, the net effect of this
paper's findings is somewhat discouraging. First, the conclusion that expected
total cost is relatively insensitive to deviations from optimal sample size
implies that less complex models, or even judgement, may be almost as satis-
factory as a formal optimization model for sample size determination. Second,
the amount of sample size reduction which may result from the incorporation
of accountant's or auditor's prior probability distributions into sample
planning appears to be quite modest, which suggests that the economic advan-
tages of a Bayesian approach may be relatively small. Third, if there are
systematic biases in the prior distribution means of accountants and auditors
(which is certainly not unlikely), then in view of the sensitivity of expected
total cost to errors in the prior mean, a Bayesian approach could have substantial
negative economic consequences. These problems certainly present a challenging
agenda for future research.
REFERENCES


American Institute of Certified Public Accountants, Codification of Statements on Auditing Standards, Numbers 1 to 26 (New York: AICPA, 1980).


Ernst & Ernst, Audit Sampling (Ernst & Ernst, 1976).


Raiffa, H. and R. Schlaifer, Applied Statistical Decision Theory (Boston: Division of Research, Graduate School of Business, Harvard University, 1961).


Decision Theoretic Estimation

Methods in Accounting and Auditing:

A Discussion

by

Soong H. Park

University of Illinois
Cushing's paper can be viewed as a feasibility study on implementation of Bayesian Decision Theory in audit practice. At the risk of offending some of my learned colleagues allow me to review some basic elements of Bayesian Decision Theory which are quite often misunderstood.

Bayesian school of statistics is different from the Classical school in its treatment of prior information. Bayesian school treats any observation as the latest addition to the pool of past observations. Therefore, empirical or judgmental prior information is formally recognized and incorporated in the analysis. Classical school is primarily interested in the information contained in the current empirical observations. Even when it allows the use of previous observations, such as pooled samples, it is accepting them as a part of the current observations.

Decision Theory (DT) is a relatively new field of statistics, the ultimate objective of which is to suggest the optimal act to take. Better understanding of the object of inquiry, which is the sole aim of the inference school, is relevant to DT only to the extent that the improved understanding leads to better action choice, thus to better results. A necessary component of the DT is the explicitly specified payoff or loss function of the decision maker.¹

¹This is not to say that the inference school does not require a loss function. In general the loss function for DT is expressed over the various act-state combination, while the inference school expresses its error measures solely on the discrepancy between the inferred and true states of nature.
Two important points need to be made clear at the outset:

1. Decision Theory and Bayesian statistics are two separate characteristics of Bayesian Decision Theory. The unique feature of DT vis a vis Inference school is the specification of act space and loss functions. Bayesian differs from Classical view in its treatment of prior information. Thus, Classical DT and Bayesian DT are both possible just as are Classical and Bayesian Inference.

2. DT and Inference school are separate and distinct framework of analysis. That is, DT is not simply an additional step of loss analysis added to the inference procedures. They are each designed for different purposes and therefore the intermediate steps are meaningful only to the extent they are evaluated with respect to the final objectives.

The major reason for the lengthy discussion above is due to the fact that Cushing's paper was advertised as a DT paper. But I do not see a decision theorist in Cushing, and as a result, this paper is at best a portrait of Bayesian Decision Theory looked through the eyes of a Classical Inferencist. This mis-match of technique and the framework resulted in awkward problem formulation and improper simplification of the problem.

Framework of Analysis

Cushing claims that the major contribution of his study is not in the development of the models for optimal sample size and terminal point estimate calculation, but in learning the sensitivities of the total cost (loss) due to deviations from the optimal solutions and errors in
input values. Since the validity of the sensitivity tests depends largely on the validity of the model used in the study, it is necessary to address the auditor's decision problem modeled in this study and the soundness of the formulated problem.

Cushing's conception of the auditor's decision problem is quite narrowly focused in that the auditor is placed at the point of "Test of Details." Kinney's [1975] representation of the auditor's decision process regarding the acceptability of an account balance is helpful to put this paper in proper perspective.

---

**Internal Control System**

- Design Evaluation → Compliance Test

**Substantive Procedures**

- Analytical Review → Decision
- Test of Details

While both Cushing and Kinney deal with the auditor's acceptance or rejection of an account balance, they differ markedly in their scope of analysis. Kinney's analysis start at the study and evaluation of internal control systems while Cushing assumes that the auditor has completed all the previous steps leading up to the direct test of balances. Thus, Kinney's analysis is to find an optimal combination of reliance on internal control systems, analytical test procedures and the extent of the direct test of balances while Cushing is concerned with the determination of optimal size of samples to examine, based on the results of the previous audit procedures.

Now, we can evaluate the appropriateness of Bayesian DT for the problem at hand. Since the auditor must combine information from various audit procedures in determining the extent of test of balance.
procedures, Bayesian school is the ideal framework for statistical sampling in auditing. To the extent that audit procedures are to provide sufficient competent evidential matter for the auditor to choose the optimal act of acceptance or rejection, DT seems to be a proper vehicle of analysis. Acceptance of Bayesian DT as the framework does not necessarily mean that the specific model used in the study is also acceptable. Remainder of the comments deal with the validity of the model used in the analysis.

We are not provided sufficient data to replicate or evaluate the accuracy of his findings. Also, the reasonableness of the parameter values used in the study has been thoroughly commented upon by the other participants at the forum. Therefore, my comments will concentrate on the methodology of the analysis, especially the simplifying assumptions made in the analysis. Specifically,

i. The assumption that the optimum sample will always yield the true "mean," $\bar{y}_{ST} = \mu$.

ii. Symmetric Loss Function Assumption.

iii. Normality Assumption: Symmetric error distribution.

Cushing assumes that, in the case of optimal decision, the sample mean obtained will always equal the population mean, thus the estimated balance will always equal the true balance. This assumption violates one of the basic principles of sampling theory. Whenever a decision maker estimates the population characteristic based on sample observations, there exists a risk of the sample yielding results that are different from the true population values, which may in turn lead to incorrect decisions. This risk is inherent in any observations less than
complete examination, and is called the sampling risk. The level of sampling risk can be controlled by the size of the sample taken by the decision maker. In audit situation, the auditor must evaluate and obtain balance between the cost of additional samples and the benefits of corresponding reduction in sampling risk. Therefore, this assumption of true mean in effect ignores the inherent sampling risk, and as a result the expected cost of the optimal decision is underestimated. Consequently, the cost of non-optimal decisions are overstated both in absolute and in relative terms. The magnitude of the errors in the analysis due to this assumption cannot be determined based on the information provided in Cushing's paper.

Cushing performed the analysis based on parameter values with some empirical support, and some values estimated so as to make the results "reasonable." He concludes that the total cost, certain sampling cost and the expected cost of potential error, is rather insensitive to the sample size and very sensitive to the non-optimal point estimate. In other words, the potential consequences of estimation error overwhelms the cost of audit procedures.

Given the importance of optimal point estimates, it is only natural to examine the assumptions violation of each may affect the optimal point estimate measures. Cushing points out that since the probability functions are assumed to be normal and the loss function symmetric, the optimal point estimate is the expected value of the distribution. The seemingly innocent use of symmetric properties in error probability distribution and the loss function resulted in hiding the impact of asymmetry of error distribution and the loss function on decisions (estimates).
First, given the symmetric probability distribution assumption, the three commonly used measures of central tendency; mean, median and mode happens to possess the same value. But when the probability distribution is not symmetric, each statistic takes on different values. Barefield [1973] noted that the expected value is the estimate that minimizes the sum of squared errors, while the median minimizes the sum of absolute value and the mode being the maximum likelihood estimator. Figure 1 displays these characteristics. Thus, the study of loss function is relevant even when the decision problem is only to obtain the "best" estimate. Real question in this case, however, is how often would an auditor encounter an asymmetric probability distribution function? This question is not addressed by the author and the answer is not readily available elsewhere either. Studies such as Neter and Loebbecke [1976] may help provide answers to this question. An important point to be remembered is that the auditor must understand which property of the central tendency measure he is using even if the numerical measures of the various estimates were identical.

Asymmetrical loss function will also cause the optimal (minimum loss) point estimate to take on values other than mean, median, or mode of the distribution. This possibility is shown on Figure 2. Obviously, when the asymmetry of both the probability distribution and the loss function is applicable, the usefulness of the expected value as the estimator is greatly reduced.
\[ f(x) = \begin{cases} kx & , \ a \leq x \leq f \\ m(h-x) & , \ f < x < h \end{cases} \]

\[ g(x) : \text{Loss Function} \]

1. \( (x - \hat{x})^2 \) : quadratic
   \[ \hat{x}^* = c \text{ mean} \]

2. \( |x - \hat{x}| \) : Linear
   \[ \hat{x}^* = d \text{ median} \]

3. \[ n, a < x < b \]
   \[ q, g < x < h \text{ Step} \]
   \[ 0, b < x < g \]

\[ \hat{x}^* = e \]

Figure 1: Asymmetric Error Distribution and the Optimal Estimates
\[ f(x) = 1 - K \cdot |x-c| \]

**g(x): Loss Function**

I. \[ |x - \hat{x}|, \ x < \hat{x} \]
   \[ 0, \ x > \hat{x} \]

II. \( \{ n, a < x < b \} \)
    \( \{ q, f < x < g \} \)
    \( \{ 0, b < x < f \} \)

**Optimal Estimate**

\( \hat{x}^* = d \)

\( \hat{x}^* = e \)

**Figure 2: Asymmetric Loss Functions and the Optimal Estimates**
To the extent that the loss function represents the auditor's assessment of the consequences of errors, it is just as natural to assess that loss in terms of utilities rather than measuring it in dollars and then translating the dollar measure into utility. If the DT model were to be used as an optimization model, the loss function must be expressed in utility measures. Once this task is completed, the maximization of expected utility is defended regardless of the shape of loss function with respect to dollar error measures. The appropriate form of the loss function can only be determined in practice. While academic research may help identify the factors that must be considered in determining the shape and parameter values of the loss functions, the ultimate player is the person taking the risk. Final assessment of the validity and usefulness of Cushing's findings must wait further developments in the area of loss function.

Summary

Earlier in the discussion, I have noted that this paper is written by a Classical Inferencist. Within DT framework, vis a vis inference framework, the terms such as maximum likelihood estimate (p. 18), lack of bias (p. 37), the amount (accountant) believes the correct value to be (p. 18) and the point estimate which are as accurate as possible (p. 49) are void of any meaning. All of these phrases are coined in the inference school. In DT the "best" estimate is the value that would lead the decision maker to the optimal decision, which may be biased and conservative. The estimates in fact do not have any inherent meaning or value, they are derived when the information is evaluated in light of the loss function and the decision situation at hand. The
familiar term "value of information" captures this very point. Again let me emphasize that once an auditor formulates the estimation problem in a decision theory framework, statistics regarding the account balances are meaningful only to the extent the measures serve as inputs to compute the expected losses of various point estimates. If the optimal point estimate turns out to be the expected value of the distribution, it is just an instance. If for some reason, the auditor is not able to pull himself away from the "expected value," then the appropriate loss function should accommodate this desire. While the simplifying assumptions by Cushing enabled him to perform the analysis with ease, the very assumptions preempted the unique decision theoretic features of the problem. His results, therefore, can only be viewed as a case study of special situations.

All of my comments are not to discourage the works such as Cushing's, rather to encourage more work in this important area. What concerns me is that the inability to express the auditor's loss functions accurately is often used as the excuse for not adopting decision theory framework. The criterion should be the appropriateness of the framework, not the availability of computational tools. Since I am convinced the auditors are in fact behaving as decision theorists, the solution is in only one approach: improve the weaknesses in the current techniques and develop better tools rather than looking the other way.
References


The Application of Regression Analysis
for Limited Review and Audit Planning
by
Abraham D. Akresh
Price Waterhouse & Co.
and
Wanda A. Wallace
University of Rochester
THE APPLICATION OF REGRESSION ANALYSIS FOR LIMITED REVIEW AND AUDIT PLANNING

by

Abraham D. Akresh, Price Waterhouse & Co.

and

Wanda A. Wallace, The Graduate School of Management, The University of Rochester Regression Consultant to Price Waterhouse & Co.

INTRODUCTION

The Symposia on Auditing Research at the University of Illinois have provided a forum for discussing the advantages of applying regression and integrated autoregressive, moving average (ARIMA) techniques to the audit. The two techniques were compared by Albrecht and McKeown (1977) at the second Symposium, and two years ago Kaplan (1979) applied ordinary least squares in building a financial planning model to produce pro forma financial statements. At that same conference a specific issue concerning regression applications, the effect of measurement error on analytical review results, was addressed by Kinney and Salamon (1979). While these papers discuss the plausibility of regression and ARIMA's application to the audit, the literature has not presented a detailed report on the actual field experience with regard to the costs and benefits of applying such statistical models.

The next logical progression in the analysis of regression applications by the auditor is to address the question:
When regression analysis is utilized

1) as part of a limited review,
2) as a means of identifying problem areas,
3) as one of the tools used to determine audit scope, and
4) as a check on the reasonableness and sufficiency of explanations for deviations,

how does it perform relative to traditional analytical review techniques?

While simulations have been presented and possible regression models using client data have been proposed in existing literature, there is no field evidence to date on the effectiveness of regression in the audit setting.

**OBJECTIVE**

This paper describes and compares the analytical review procedures traditionally applied with the application of regression analysis. As a first step, the technical validity of the regression models specified by the auditor for a limited review is analyzed. Then the regression predictions are scrutinized to determine whether the regression tool has missed problem areas, provided additional insights, or simply generated random noise. Outliers have been investigated where warranted for limited review purposes and the significance of the information provided by the regression outliers to the limited review and subsequent audit engagement is considered.

To further measure the relative performance and efficiency of traditional analytical review techniques and regression, differences in audit staff time are considered. Training requirements and computer and related direct costs are also considered. Areas of the client's operations that are subjected to further review work are compared. In an attempt to assess the auditor's
model-building capabilities, a stepwise (optimum) regression (see Maddala, 1977) program for model selection is applied to all available data related to each account of primary interest, and the results are compared to the auditor-selected model. Finally, the effects of the two modes of analytical review work on audit planning are assessed. Practical problems in applying regression analysis and relevant factors in determining where and when to apply regression are described based on the field application.

FIELD EXPERIMENT

A chronological summary of the research approach to the regression field experiment is presented in Figure 1. In this field experiment the client is a public utility with gas and electric operations. The utility is considered to have a good system of internal accounting control, including reliable monthly financial, production, and customer data. Budgeted data are also available on a monthly basis.

In order to assure a "controlled" experiment, the audit team first performed and documented traditional analytical review procedures and conclusions drawn. Then the audit manager for the limited review engagement selected the accounts (dependent variables) to be analyzed by regression based on the key facets of client operations that would be of concern in any limited review engagement. The manager also specified the descriptor (independent) variables for each model based on his knowledge of client operations and regression analysis as an audit tool (having attended a 1 1/2 day training session on regression as well as all firm training on statistical sampling). The audit team agreed on the desired precision for the variables being described by the regression; this desired precision is 1/2 the monthly materiality cut-off per account which would be of concern in an audit. It was acknowledged that the
Chronological Summary of Research Approach to the Regression Field Experiment

Met with Audit Staff to
(a) Gain familiarity with client operations
(b) Briefly introduce the regression concept
(c) Direct the preparation of machine readable data

Reviewed Client's Public Reports and Working Papers

Tested available time-sharing packages for potential adaptation to an audit setting

Met with Client to describe the regression concept and to provide a question/answer session for client personnel

Performed a regression analysis on an available time-sharing system and, upon review of existing problems, decided to develop own software package

Ongoing revision and testing of software

Programmed and tested software

Prepared materials for and Conducted a Regression Training Session for 1 1/2 days for 28+ managers

Auditor who attended Training Session selected accounts to be analyzed by regression for the limited review and identified models, including
(a) expected signs of the regression coefficients
(b) desired materiality cut-offs

Ongoing revision and testing of software
Performed a Limited Review of Quarter 1 & 2 Without Regression

Performed the Regression Analysis for accounts selected for the Limited Review application

Audit Manager, Senior, and authors compared Limited Review regression results with and without the regression application and additional investigation was performed where suggested by the comparison

Integrated the Limited Review work in the Audit Planning Process (on a Complementary basis only)

Plan to perform 3rd Quarter Limited Review, relying primarily on regression

Verified regression output through:
(a) detailed analysis of programming
(b) parallel analysis through another time-sharing system (where practical)

In Process

Currently performing other field tests on limited review, audit planning, and year-end audit applications using regression analysis
limited review would not have to reach the audit precision requirement because
the auditor's responsibilities are less and because the nature of procedures are
such that they are a review rather than an audit. However, since the limited
review, rather than being an audit of financial statements for the quarter, is
a part of the continuing audit, and since the review results are to be used for
audit planning, the benchmark used in this field experiment to evaluate the
regression findings is that precision required for an audit engagement.

THE AUDITOR-SPECIFIED MODELS

The Regression Training Session for audit managers emphasized the
availability of numerous approaches to model-building, each of which reflects
a different audit setting.

1. External Data

There are two major categories of external data available to the auditor
for regression model-building: (1) data available outside of the firm:
e.g., macroeconomic statistics like the Consumer Price Index (CPI), industry
averages, borrowing rates, temperature statistics, and customer base and
(2) data generated by the firm external to the accounting function, e.g.,
production and sales statistics and information concerning existing plant
capacity. The use of external data is in the auditor's interests in that
the audittee does not have control over category (1) and category (2) pre-
sumably would not be subject to manipulation by those client employees gen-
erating accounting records. Hence, the strength of evidence generated from
such "outside" data would be greater than the strength of evidence generated
solely from accounting records.

2. Budgeted Data

Although budgets are typically distributed by the accounting department,
they are formulated and subject to review by line management. The "inde-
ependent" (in time and possibly function) budget process can provide external/ internal evidence as to the reasonableness of actual accounting records. Traditionally auditors compare budgets to actual performance as part of analytical review work as a means of identifying variations from expected performance. However, it is possible to expand this "balance" check to a detailed analysis of account interrelationships by including budgeted data in regression model building.

3. Internal Data

The use of external data in prediction models is the regression application least open to criticism from the evidential/liability perspective of the auditor. However, it is well recognized that the audit function utilizes the full spectrum of available evidence, both externally and internally generated, as a basis for formulating a judgment concerning the fairness of financial statement presentation. An important determinant of how much reliance can be placed on a client's internal records is the quality of the internal control system. When the controls are to be relied upon by the auditor in performing substantive tests, it is appropriate to reflect such reliance in the regression tool (i.e., by utilizing internal data). Internal data may be audited or unaudited in nature. By using lagged models, i.e., predicting this year's sales from last year's, the auditor can avoid relying on the current period's unaudited data in formulating forecasts. However, it must be acknowledged that if monthly data is used, it is not valid to call the prior period's monthly figures audited, despite the audited status of the annual financial figures. If the same controls exist for the current period as prior, and if these controls are deemed adequate over monthly financial statements based on a current review and compliance test of the system, a basis exists for utilizing current
period unaudited data in model-building. Even in the absence of good internal control, internal data can be utilized in regression models, given the audit plan requires testing of the account used via a separate audit technique prior to reliance on the model's output.

Among plausible internal data models are general accounting-based models with no company specific adjustment. To some degree, this approach is demonstrated by Kaplan (1979) in his development of a financial planning model.

4. Integrating Internal and External Data

Knowledge of the internal control system of a client provides the auditor with an assessment of how likely particular information in the accounting and management information systems will be in error or subject to manipulation. This concept of risk exposure provides a basis for formulating regression models that utilize internal data. For example, if segregation of duties is inadequate over cash and accounts receivable record-keeping, a regression model of cash based on receivables would be extremely limited in its audit implications. Given the extraction of cash assets can be hidden via manipulation of receivable records, and both the dependent and independent variable are related to the quantity of total collections and errors reflected therein, trouble-spots across these two accounts would not be detected by applying such a model.

a. Endogeneity

Endogenous variables are those determined within the economic system whereas exogenous variables are given from outside the economic system. Typically a dependent variable is endogenous while independent variables are assumed to be exogenous. However, practically speaking, variables determined by an economic system are classified as exogenous if not subject to control by the same party determining the dependent variable. Hence, in econometric terms, the concern
for endogeneity in model-building is analogous to an auditor's concern over internal control. It can easily be argued that all accounting numbers are endogenous to the accounting system of the client and, in fact, in a broad sense that almost all variables are endogenous. However, the primary concern is whether something on the right side of the regression equation is "masking" a change in the dependent variable. In other words, is something random influencing both the error term of the independent variable and the error term of the dependent variable? In auditing terms, the key question becomes: can the relation of the independent variables to the residual term operate to systematically "cover-up" a defalcation or accounting error? A careful definition of endogenous—as control by those generating the dependent variables over the independent variable(s) to the extent that trouble-spots could be hidden—permits the auditor to formulate reduced-form equations with only exogenous variables on the right-hand side of the equation.

b. Simultaneous Equations

An alternative approach to endogeneity and the simultaneous determination of two or more variables would be the development of a simultaneous equation system (see Kaplan, 1979). Basically, such an approach quantifies and permits adjustment of the correlation between explanatory variables and the error term. However, the development of a simultaneous equation system asserts a clear theory of the relationships between multiple equations. While mathematical relationships between accounts are known via the double-entry accounting system, there is no complete theory of the firm, or the relation of accounting to such a theory, that permits the full modeling of production, investments, and other operations of a client for use as a basis for developing a system of equations. Since the correction factor estimated via simultaneous techniques is only as
good as the underlying model, there are clear advantages to developing reduced-form equations that reflect the specific concerns of the auditor by defining endogeneity in relation to individual accounts.

c. Trade-offs in Precision

While the risk inherent in the use of internal data has been emphasized, offsetting benefits are derived from integrating internal and external data in regression model building. Regression models' resolving power can be improved by incorporating that data to which the auditor has unique access when formulating estimates of what book values should be for the period under audit, i.e., internal data. Traditionally analysts and other external model-builders have been forced to use external quarterly or annual data for model construction, whereas auditors will have access to monthly accounting statistics.

All of the available budgeted and actual monthly account balances for the Statement of Financial Position and for the Statement of Earnings from January, 1973 through June, 1980 were put into machine readable form for use in the first Limited Review and in subsequent engagements. The data base from January, 1973 through December, 1979, or a total of 84 observations, is available for constructing a base model for use in predicting account balances for the first quarter's review. This base period was utilized to compare alternative approaches to model building.

Table 1 exemplifies the trade-offs in model precision which accompany the use of external versus internal data. As expected, the combination of internal and external data, both production and macroeconomic statistics, increases regression models' resolving power. From the high of $30 million to a low of $3.6 million for the model precision of electric revenue, the trade-off is clearly a substantive amount. Although the $R^2$ shift is as
### TABLE 1

**Alternative Approaches to Model Building**

From an Auditor's Perspective:

Observed Tradeoffs in Precision

<table>
<thead>
<tr>
<th>Model</th>
<th>Precision</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Standard Error</td>
<td></td>
</tr>
<tr>
<td>External Data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Electric Revenue: Degree Days &amp; CPI</td>
<td>7,430,360</td>
<td>.956 (.955)</td>
</tr>
<tr>
<td>- Gas Revenue: Degree Days &amp; CPI</td>
<td>8,251,992</td>
<td>.862 (.854)</td>
</tr>
<tr>
<td>External or Easily Verified Data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Gas Revenue: Degree Days &amp; CPI &amp; Number of Gas Customers</td>
<td>6,744,360</td>
<td>.911 (.902)</td>
</tr>
<tr>
<td>Production Statistics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Electric Revenue: KWH's-Residential</td>
<td>30,105,520</td>
<td>.263 (.254)</td>
</tr>
<tr>
<td>- Gas Revenue: MCF's-Residential</td>
<td>7,005,432</td>
<td>.898 (.895)</td>
</tr>
<tr>
<td>Budget Comparisons</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Electric Revenue: Budgeted Electric Revenue</td>
<td>10,868,240</td>
<td>.907 (.906)</td>
</tr>
<tr>
<td>- Gas Revenue: Budgeted Gas Revenue</td>
<td>5,108,152</td>
<td>.946 (.944)</td>
</tr>
<tr>
<td>Combined Accounting, Production, and External Data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Electric Revenue: Degree Days, CPI, KWH's-Residential, Budgeted Electric Revenue, Production Expenses-Electric - Gas Revenue: Degree Days, CPI, Number of Gas Customers, MCF's-Residential, Budgeted Gas Revenue, Production Expenses-Gas</td>
<td>3,582,710</td>
<td>.990 (.989)</td>
</tr>
<tr>
<td></td>
<td>1,067,520</td>
<td>.998 (.998)</td>
</tr>
</tbody>
</table>

---

A Replication of Selected General Accounting Based Models Reported in Kaplan (1979)

<table>
<thead>
<tr>
<th>Model</th>
<th>Precision</th>
<th>( R^2 )</th>
<th>Durbin-Watson</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kaplan—Cost of Goods Sold: Sales</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Replication—Production Expense—Electric: Electric Revenue</td>
<td>--</td>
<td>.88</td>
<td>--</td>
</tr>
<tr>
<td>Auditor-Specified Model—Production Expense—Electric: Electric Revenue, Fuel CPI, N/I, EPS-CS, KWH—Residential, Degree Days</td>
<td>4,083,668</td>
<td>.96</td>
<td>1.076 P</td>
</tr>
<tr>
<td>Replication—Production Expense—Gas: Gas Revenue</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Auditor-Specified Model—Production Expense—Gas: Gas Revenue, CPI, N/I, EPS-CS, MCF’s—Residential, Degree Days</td>
<td>3,283.9</td>
<td>.95</td>
<td>1.99</td>
</tr>
<tr>
<td>Replication—Production Expense—Gas: Gas Revenue</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Auditor-Specified Model—Production Expense—Gas: Gas Revenue, CPI, N/I, EPS-CS, MCF’s—Residential, Degree Days</td>
<td>2,983,708</td>
<td>.96</td>
<td>.41 P</td>
</tr>
<tr>
<td>Kaplan—Depreciation: Gross Building, Machinery &amp; Equipment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Replication—Depreciation—Electric Plant</td>
<td>--</td>
<td>.60</td>
<td>--</td>
</tr>
<tr>
<td>Auditor-Specified Model—Depreciation—Electric Plant, Electric Revenue, Gas and Electric CPI, Budgeted Depreciation, Accumulated Depreciation Levels Model</td>
<td>139,763.9</td>
<td>.924</td>
<td>1.03 P</td>
</tr>
<tr>
<td>First Differences</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>120,751</td>
<td>.944</td>
<td>1.88</td>
<td></td>
</tr>
<tr>
<td>143,522</td>
<td>.192</td>
<td>2.46 II</td>
<td></td>
</tr>
</tbody>
</table>

**P** = Positively Autocorrelated at a .05 level of significance

**I** = Inconclusive with respect to positive autocorrelation

**II** = Inconclusive with respect to negative autocorrelation
marginal as .96 to .99 for an alternative model, the accompanying precision drops in half when external data is augmented by accounting internal data.

Due to the excellent client controls and recent review and test of the client's entire information system, the auditor chose to integrate internal and external data in specifying regression models. Whenever possible, "uncontrollable" descriptor variables are included in the regression models as means of complicating, if not precluding, any attempts at disguising actual financial performance. The Appendix, Tables A, B, C, D, E, F and G present the auditor-specified models and the auditor's desired precision for the seven accounts selected by the audit manager to be analyzed using the regression tool.

The literature suggests that the development of classes of regression models, such as general accounting-based models could be an effective means of lowering the costs of model specification without a substantial loss in the resolving power of regression models. To provide some measure of the trade-off of such a general model approach, Table 1 compares the model precision of three general relationships suggested by Kaplan (1979) with the precision obtained by the auditor-specified models. The achieved $R^2$ for the utility client is similar to that achieved by Kaplan when applied to a firm in a different industry; these values suggest substantial explanatory power in the levels form of the regression models. Further, the model precision, while consistently looser than that achievable through specific client-level model building, is encouraging as an indicator that even generalized models can have reasonable resolving power.

Obviously, however, such general models must be fully tested regarding least squares assumptions, and subjected to the real test of predictions
from accounting-based simplistic models: do the models give false signals or do they represent audit trouble spots? These tests are required as are numerous replications to provide complete answers as to the trade-off's in precision of tailoring a regression model versus applying a generalized model. However, such scrutiny is beyond the scope of this paper. The field experiment is directed at studying auditor-specified models at the client level. Having selected the regression relationships of interest, the next task is to access a time-sharing system and estimate the models.

**Regression Software**

As Figure 1 indicates, existing time-sharing packages were considered for potential adaptation to an audit setting. However,

- the excessive output form in econometric terms,
- the inefficient manipulation of future data for repeated forecasting,
- the "backward" approach to reporting outliers by considering book values as true values rather than the regression estimates (i.e., an incorrect formulation of confidence intervals relative to the auditor's concerns),
- the lack of control over the "black box" which statistically checks a proposed regression model, due to the limited accessibility to program coding permitted by companies supplying time-sharing packages,
- the auditor's concern for the technical validity of regression in light of known problems with accounting time series data [see Benston (1966), Comiskey (1966), Jensen (1967), Deakin and Granof (1974), and Kinney and Bailey (1976)] for which statistical checks in existing time-sharing packages are incomplete, and
- the training cost and regression utilization time savings which could accrue from a tailored time-sharing program which automatically performed statistical checks, directed the auditor in selecting appropriate data transformations to "fix" the deficiencies identified, and warned the auditor concerning model limitations, in a simplified output form which utilized the terminology already understood by auditors familiar with statistical sampling,
provided the basis for the decision by Price Waterhouse & Co. to develop its own software package.

Figure 2 provides an overview of the software design specifications to date. Since the software development is still in process, the regression application reported in this paper was generated using both this newly developed program and an existing time-sharing package for general statistical analysis. Through this parallel analysis, the propriety of the new package's regression estimates was verified and most of the statistical checks not yet implemented in the new software were able to be generated. Due to the authors' familiarity with regression model-building, some of the statistical checks performed in this study were via examination of plots which requires skill in model-building beyond that which will be required of auditors, since alternative formula checks will be implemented in the software.

**STATISTICAL CHECKS**

The Appendix reports the estimated regression models, including test statistics for numerous technical validity checks. Models are reported for 36 and 84 observations to provide evidence as to the trade-off between the number of observations and the stability of proposed regression models, as reflected by the achieved model precision. The literature presently endorses model-building in the audit setting with 36 monthly observations (see Stringer, 1975; Deakin and Granof, 1974; Kinney, 1978), and this study explicitly tests the propriety of that endorsement. Models are also reported in the Appendix in both levels and first-differenced form. The advantages of estimating models with changes, i.e., first differences, are developed by Granger and Newbold (1974) and Plosser and Schwert (1978).
FIGURE 2
Overview of Software Design Specifications

START

BASE MODEL

a. Compute first order autocorrelation

b. Run Levels Model (L)

c. Select Model Form: L, FD, or CO with lowest standard error

Is value ≥ 0.87?

Yes

Run First Difference Model (FD)

NO

Run Cochrane-Orcutt (CO) Model

STATISTICAL CHECKS

Nonlinearity or Autocorrelation
- Runs Test
- Contingency Table
- Durbin-Watson

If any are significant

A

Heteroscedasticity or Non-Constancy of Residual Variance
- Nonparametric rank correlation of (residual) and value of descriptor (independent) variable(s)
- Goldfeld and Quandt

If any are significant

B

Normality of Errors
- Kolmogorov-Smirnov test
- Moment check
- Shapiro-Wilk statistic

If any are significant

C

- Internal Checks:
  - Significant digits made consistent across variables
  - Base Model observations reported, with output as to data file observations not in use

- Reminders Provided:
  - Warnings as to effects of statistical check findings
  - Warnings to check coefficients' direction and magnitude
  - Warnings as to allowable number of seasonal dummies permitted

- Options Provided:
  - Transformations: Add, Subtract, Multiply, Divide, Lag, Log
  - Creation of Dummy Variables: Seasonal, Linear Trend, Creation from Values
  - Stepwise: Forward selection and backward elimination with an optimum regression test on the number of regressors (descriptor or independent variables -- 's) selected.

- Possible Additions to Statistical Checks:
  - Outlier checks: Single Row Diagnostics and Wilks' A
  - Multicollinearity checks: Regression Coefficient Variance Decomposition combined with Single-Row Diagnostics
FIGURE 2 (continued)

Recalculate Statistical Checks

a. Log of X
b. If still significant log of X and y

Is a multiplicative relationship suspected?

b

Divide by the X variable(s) and recompute the model

Take log of y

If still significant

Recalculate Statistical Checks

Select FD model form

If still significant or if already FD or CD

Recalculate Statistical Checks

WARNING

D

If still significant

Recalculate Statistical Checks

WARNING

E

If still significant

Recalculate Statistical Checks

WARNING

F

Take log of y

If still significant

Recalculate Statistical Checks

WARNING

F

If still significant

Recalculate Statistical Checks

WARNING

F
FIGURE 2 (continued):

Proceed regardless of results, but output the quantities for the statistical checks on the model form selected, based on all the above tests.

OTHER CHECKS OF BASE MODEL

Spurious relationship concerns
- $R^2 \approx 0$
- $\eta^2, \tilde{R}^2$ for First Difference Model
- Standard Error

Multicollinearity concerns
- Correlation Matrix > .8 values
- Halvovskiy statistic
- Are all $B_i$ insignificant at .05 level

Seasonality
- Autocorrelation values for Lag 1 through 12

Model Shifts
- Chow test within Base Period
- Comparison of Base Period with Audit Period
- Comparison of Standard Error for last 36 obs. versus total base

IF $R^2 \approx 0$
STOP
Otherwise output results

Consider seasonal dummies, differencing, or use of lagged values

Consider deleting variables. WARNING

IF < 24 obs., available w/o shift STOP
Otherwise output results

PREDICTION PHASE

Simultaneous Prediction intervals at a 95% Level are computed and the amount by which specific monthly observations lie outside lower or upper bounds is reported. Tests will be conducted regarding the:

1. Preferred confidence level to apply in identifying outliers,
2. Savings from applying Kinney's (1979) algorithm to standardized residuals with and without use of the prediction interval as a filter,
3. Advantages to adapting the monthly approach to an annual risk level approach.

Computation of 1/2 the 95% Prediction Interval as a Measure of Model Precision per prediction, for Comparison to the Auditor's Desired Precision.
Basically, the differenced model, when compared with a levels regression assists the model-builder in analyzing a model's specification and in avoiding "nonsense" regressions involving autocorrelated series.

The detailed results of such comparisons and numerous additional model checks are summarized in Table 2. Basically the table reaffirms the statistical issues in model-building which can be expected to occur in audit regression applications. The frequency with which autocorrelation and model shifts occur, multicollinearity is confronted, and transformations are required for selected regression models is reported. Nonlinearity, the lack of normality, and heteroscedasticity are not problems in the data base under study. The seasonality of the data is apparently controlled by descriptor variables related to the weather and to production statistics. Autocorrelation is typically controlled via differencing or the change in model size as a result of adapting to the results of the test for a model shift.

While the 36-observation models dominate, there are cases, such as the electric revenue relationship, in which model precision is improved by utilizing all 84 observations in the data base. Spurious relationships are not a problem, although the depreciation model specified by the auditor lacks substantial real explanatory power. While multicollinearity is present, it does not appear to substantially harm the models, i.e., the linear combinations of estimated regression coefficients appear to be well-determined. However, the commonality of high correlations among accounting time series suggests that whenever a selected variable has an insignificant t-value, a nonzero probability exists that multicollinearity is present which, if harmful, can cause an understatement of the variance of the model estimates. Although in this
### Table 2

Results of Statistical Checks on Base Models

<table>
<thead>
<tr>
<th>Issue Under Study</th>
<th>Total Number of Models Examined</th>
<th>Number of Significant Findings (Significance Level)</th>
<th>Persistent Problems in Selected Models Pre-Transformation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Autocorrelation</strong></td>
<td>28</td>
<td>5 ( I ) 6 ( I I ) 3 ( P ) 6 ( N ) (.05)</td>
<td>4 ( I I ) (.05)</td>
</tr>
<tr>
<td>- Durbin-Watson</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Runs Test</td>
<td></td>
<td>11 (.05)</td>
<td></td>
</tr>
<tr>
<td><strong>Normality</strong></td>
<td>28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Kolmogorov-Smirnov Statistic</td>
<td></td>
<td>1 (.05); 11 (.01)</td>
<td>1 (.05); 1 (.01)</td>
</tr>
<tr>
<td>- Skewness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( x &lt; -1 )</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(-1 &lt; x &lt; -.5 )</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( .5 &lt; x &lt; 1 )</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( 1 &lt; x )</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Kurtosis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( x &lt; 2 )</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( 4 &lt; x &lt; 6 )</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( 6 &lt; x &lt; 10 )</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( 10 &lt; x )</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Nonlinearity</strong></td>
<td>7 (17 plots)</td>
<td>2</td>
<td>NL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Plots of each variable to be explained with each significant descriptor variable were examined</td>
<td></td>
</tr>
<tr>
<td><strong>Model Shift Chow Test</strong></td>
<td>[1st 48 observations compared with last 36 observations]</td>
<td>7</td>
<td>2 (.05) 2 (.01) Selected 36 obs. Model</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Spurious Relationship</strong></td>
<td></td>
<td>28</td>
<td>1 (.05) 1 (.09) 1 (.10) -</td>
</tr>
<tr>
<td>- Significance of F statistic other than .01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Standard error of First Difference (FD) Model relative to Levels (L) Model</td>
<td>28</td>
<td>1:25.7 1:29.6 1:29.9 2</td>
<td>1:36 1:34.8 1</td>
</tr>
<tr>
<td>25 to 30% higher</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; 30% higher</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
TABLE 2 (continued)

-Substantial Decline in $R^2$ for FD relative to L Model

-Number of Descriptor (Independent) Variables with less than a 50% significance
  1 Descriptor: frequency
  2 Descriptors: frequency

-Frequency of observed differences between the expected and actual sign or direction of a significant coefficient

7 (48 var.)

■ Heteroscedasticity
-Plot of residuals vs descriptor variables

■ Multicollinearity
-All $\beta_i$ insignificant $\theta .05$
-Correlation between two descriptor variables
  (Quantity: Frequency) $.8 < \chi < .9$
  $.9 < \chi$

■ Seasonality
-Autocorrelation at lags 1 through 12
  Number of lags with values $>|.5|$: frequency
  (a) Variable to be explained
     (dependent or $y$ variable)
  (b) residuals
-Seasonal Dummies with $>.2$ correlation with $y$ variable (Quantity: Frequency)

- Footnotes are provided on next page.
Footnotes for Table 2

1. Inconclusive as to Positive Autocorrelation @ 95% (.05 significance)

2. Inconclusive as to Negative Autocorrelation @ 95%

P = Positive Autocorrelation @ 95%

M = Negative Autocorrelation @ 95%

I. When transformed to the log of y being explained by the log of the X's, the plots of residuals on normality probability paper suggest the normality concerns were corrected. In fact, the Kolmogorov-Smirnov statistic for AFDC dropped to .0823, not significant at a .05 level. Note that the 36 levels model for depreciation was transformed; since logs of negative differences preclude the selected model's transformation.

A plot of residuals on normal probability paper also indicates a departure from normality.

NL. The AFDC plot versus interest Long-Term Debt indicates slight nonlinearity, taking the general form (a). The Interest L-T Debt plot versus utility bond rates also indicates slight nonlinearity, taking the general form (b).

M = One variable of each descriptor pair which was correlated more than .8 was deleted from the auditor-specified models. The newly estimated models were compared to the full models along the following attributes:

<table>
<thead>
<tr>
<th>Account Being Explained</th>
<th>Descriptor Variables Explained</th>
<th>Changes in Coefficients' Significance</th>
<th>Changes in Coefficients' Direction (Is this coeff. signif.)</th>
<th>Model Precision/Changes in Outliers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electric Rev.</td>
<td>Degree Days, CPI</td>
<td>KWH's--Res., t = 5.6</td>
<td>Constant (No)</td>
<td>4068/None</td>
</tr>
<tr>
<td>Prod. Exp.--Electric</td>
<td>Degree Days</td>
<td>Nu of Gas Cust., t = 2.1</td>
<td>None</td>
<td>[Runs Signif.: -2 Ratio]</td>
</tr>
<tr>
<td>Prod. Exp.--Gas</td>
<td>Fuel CPI</td>
<td>Constant, t = -2.3</td>
<td>None</td>
<td>3761/None</td>
</tr>
<tr>
<td>Depreciation</td>
<td>KWH's--Res.</td>
<td>Degree Days, t = -1.8</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td></td>
<td>EPS--CS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Net Income</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MCF's--Res.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Elec. Rev.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gas &amp; E1. CPI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Constr. WIP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AFDC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interest--LTD</td>
<td>CPI, AFDC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(runs @ .05)</td>
<td>Div.--P.S.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The shift in coefficient significance suggests multicollinearity which affected model estimation was present; however, the directional sign of the coefficients was stable, and the model precisions very close to the full model, with the exception of Prod. Exp.--Gas.

**For example, degree days and seasonal dummies are correlated.

S = A stepwise regression model for gas revenue including all 11 seasonal dummies and a linear trend variable yielded the eighth dummy and the linear trend variables as significant at a .05 level, dropping number of customers, CPI, and budgeted gas revenue. The model precision was 928.8, comparing favorably to the 1250 desired precision. Similarly, a test of production expenses--gas led to a 1.7 t-value for season 1 and a model precision of 736.63, comparing favorably to the 800 desired precision. CPI and net income were dropped.
paper multicollinearity was examined by deleting variables, there are risks
to the specification of the model when such action is taken and alternative
techniques for identifying when collinearity exists, its magnitude, and
whether the collinearity is just degrading (i.e., regression coefficients
could be improved by adding better conditioned data) or is harmful due
to inflated variances are being explored for use in the audit software (see
Belsley, Kuh, and Welsch, 1980).

The regression model selected per account, as a result of the statistical
evaluation of each model form, is reported in the Appendix. From a perusal
of the far right column of Table 2, it is apparent that changes in model
size and the use of first differences result in a set of technically valid
base models for use in the application of regression analysis for limited
review and audit planning.

**FLUCTUATIONS IDENTIFIED USING TRADITIONAL ANALYTICAL
REVIEW PROCEDURES: DOES REGRESSION ANALYSIS
PROVIDE SIMILAR INFORMATION?**

Traditional analytical review procedures currently followed for limited
review engagements focus on variation analysis. Auditors are instructed to

Be alert for interrelationships of elements of
financial information that would be expected to
conform to a predictable pattern based on the
company's experience. (Source: Audit Program,
Price Waterhouse & Co.)

Table 3 summarizes the three approaches to variation analysis applied in the
limited review engagement and the explanations for variations identified
through these traditional procedures for the seven accounts later subjected
to regression analysis.
<table>
<thead>
<tr>
<th>Account Under Analysis</th>
<th>Explanation Expected to be Captured by Auditor-Specified Regression Models</th>
<th>Additional Descriptor Variables Suggested by &quot;Explanations&quot; to Capture &quot;Causes&quot; of Observed Outliers</th>
<th>Original Number of Outliers for all Model Forms</th>
<th>Explanation Related to Outlier and Not Captured by Model</th>
</tr>
</thead>
</table>
| Electric Revenue       | a. fuel adjustment clause  
  b. rate increase in March 1980  
  c. offset: drop in KWH's | Not a significant variation  
  1. a) Fuel CPI  
  2. b) Fuel CPI  
  2. c) KWH's--other | 3  
  - Degree Days Dropped (Insignificant)  
  - Fuel CPI substituted for CPI (2; 3426) | February: Strong conservation efforts were not as prevalent in prior years; revenues are immediately preceding a rate increase (Also, see footnote ** to March--1.b; 2.b) |
| Gas Revenue            | a. Purchase gas adjustment clause  
  b. increase in charges for purchased gas  
  c. insignificant rate increase in March 1980  
  d. gas conversions increased MET's  
  e. change in customer mix | Not a significant variation  
  1. a) Gas and Electric CPI  
  1. e) MET's--other  
  1. e) # of Customers--other | 1  
  - Added MET's--other  
  and # of Customers--other  
  - Gas and Electric CPI substituted for CPI (0; 932) | N/A |
| Production Expenses-Electric | a. average cost of coal and oil increased  
  b. KWH's decreased with outages requiring higher purchases of elec. | a. additional purchase of power due to outage and lack of expected commercial use of a plant  
  1. a) Gas and Electric CPI  
  1. b) KWH's--other  
  2. c) Gas and Electric CPI  
  2. d) KWH's--other  
  3. a) Gas and Electric CPI and KWH's--other | 1  
  - Added Gas and Electric CPI and KWH's--other (0; 3167)* | N/A |
| Account Under Analysis | (1) Qtr. 1, 1980 compared with Qtr. 1, 1979 | (2) Qtr. 1, 1980 compared with Budgeted Qtr. 1, 1980 | (3) Qtr. 1, 1980 compared with Budgeted Qtr. 1, 1980 | Explanation Expected to be Captured by Auditor-Specified Regression Models | Additional Descriptor Variables Suggested by "Explanations" to Capture "Causes" of Observed Outliers | Original Number of Outliers for all Model Forms | Model Changes
(Number of Outliers After Addition of Suggested Descriptor Variables; Model Precision for New Model) | Explanation Related to Outlier and Not Captured by Model

Production Expenses--Gas

- a. Increase in cost of purchased gas
  - a. Seasonal effects increased demand
  - b. Cost of purchased gas increased

Depreciation

- a. Utility plant in service increased
  - a. Increase consistent with increase in utility plant in service

Allowance for Borrowed Funds Used During Construction (AFDC)

- a. Increase in construction work-in-process (CWIP) and minimal effect of increase in Nuclear Fuel in Process (NF)
  - b. Conductive weather relative to 1979
  - c. Increase in AFDC rate in Nov., 1979 and increase in Jan., 1980

Interest: Long-term Debt

- a. More debt Dec., 1979

**A runs test is significant at 95% (.05 level) with a ratio of -3, for production expense--electric.

**Besides, the sum of 1/2 the prediction interval and the amount by which the outlier is below the lower bound is less than the auditor's desired precision.
TABLE 3 (continued)

Note: A computation of the annual growth rate implied by a change in an account balance from one month to another would have disclosed:

<table>
<thead>
<tr>
<th>Description</th>
<th>Comparison</th>
<th>Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electric Revenue relative to</td>
<td>2/80 to 3/80</td>
<td>246.1%</td>
</tr>
<tr>
<td>Production Exp.--Electric relative to</td>
<td>2/80 to 3/80</td>
<td>965.2%</td>
</tr>
<tr>
<td>Production Exp.--Gas relative to</td>
<td>2/79 to 3/79</td>
<td>324.8%</td>
</tr>
</tbody>
</table>

As outliers, similar to the traditional and regression techniques. However, while such an approach can highlight account problems, it creates the same demand for investigation and the lack of accountability of the changes from known factors as a simpler percent variation analysis.
When the audit manager selected the descriptor variables for the seven regression models he was unaware of the findings of the traditional review procedures. Yet, the auditor-specified models contain variables which can be expected to capture the effects of several of the explanations provided in the working papers for observed fluctuations in the accounts of interest, as noted in Table 3.

One of the key advantages of applying regression analysis as an audit tool—commonly cited in the literature—is that it provides a more disciplined approach to analytical review. Not only is the auditor forced to think about what variables are related to what other variables, (s)he is able to test the reasonableness of the amount of fluctuations in account balances.

The presently used nonstatistical analytical review techniques can best be viewed as confirming the direction of change, rather than even the approximate amount of change. For example, assume sales went up by 15% over the same period of the prior year. The auditor will ask the client why, and the client might say "because we had a colder winter," "because we sold more kilowatt hours (KWH's)," "because we have more customers," and "because we raised our prices to keep up with inflation." The auditor will test to see that these appear true, i.e., that there are more degree days, that more electricity was sold, that there are more customers, that prices were raised for inflation and so forth. This approach to a limited review is reflected in Table 3. However, with this approach, the auditor does not know whether the explanations provided should have accounted for a 5% increase or a 50% increase, or the actual 15% increase. Also, (s)he does not know if there are counterbalancing factors which should have caused sales to
go down. All (s)he knows is that (s)he has identified factors which should cause sales to go up.

Thus, the main advantage of the regression model is that it helps the auditor test the reasonableness, not only of the direction of change, but of the amount of change. With regression, the auditor has satisfaction that sales should have gone up about 15%, not 5% or 50%. Beyond formulating an estimate of book value, the regression tool provides a means of assessing the reasonableness of client explanations for significant fluctuations. By formalizing the decision process for the acceptance of client explanations, the objectivity of the auditor's evidential base is improved.

Table 3 reports those additional variables which are suggested for model inclusion by the explanations of the client, and the effect on observed outliers of including such variables in the regression model. Four of the original ten outliers are explained by the expanded models. A fifth outlier is explained by a rate case for which no variable was included in the regression evaluation.

As the investigation of regression outliers proceeded, the long-range perspective provided by the regression technique relative to traditional review procedures became apparent. For example, the one-year comparison of interest expenses on long-term debt, performed in the limited review engagement, detected an increase in interest expense and related that increase to a new issue, netted with two bond retirements. However, the regression models in levels form indicated that the expenses were less than expected from a historical perspective. Further investigation indicated that the typical management financing policy had shifted and that anticipated debt
issues had not been made. Such information on the financing policies of a
client relative to the operating performance (note the inclusion of a Net
Income descriptor variable) and expansion practices (note the inclusion of
the account Allowance for Borrowed Funds Used During Construction in the
regression model) of the company is certainly more relevant information to
an auditor than a simple one-year change in account analysis. Hence, the
power of regression in revealing policy shifts is supported by this field
experiment and provides one example of an insight gained from the regression
application which was not gained from the application of traditional analytical
review procedures. This insight corresponds to a sixth and seventh outlier.

An eighth outlier was attributed to the greater conservation efforts
which occurred in 1980 from a historical perspective, as well as the month's
relationship to a rate case. The rate case and its effects on surrounding
months might be captured by a qualitative dummy variable which utilizes a
one value for each month that a rate case occurs and a zero value for the
remaining (84 less the number of rate cases) observations, though the
variable was not constructed for testing in this field application.

The ninth and tenth outliers remaining were present only in the 84-observation
levels form of the regression specification for gas production expenses. Although
the Chow test for model shift is insignificant for this model, there were signifi-
cant (at a .05 level) runs detected in the residuals. Although the Cochrane-
Orcutt technique may have adjusted for these runs in the 84-observation model,
in this field experiment the 36-observation model was selected for use in the
limited review due to the absence of statistical complications, and no outliers
were indicated. It is possible that the 84-observation model results can be
attributed to the statistical problem, or, perhaps to the inability of the
auditor-specified model to capture the change in the cost of purchased gas. Rather than a gas and electric CPI, a variable of price lists for gas purchases from 1973 to 1980 would have more power in detecting the upward movement of production expenses—gas. The outliers were not subjected to further investigation for limited review purposes, although it is likely some additional model development would be appropriate for an audit engagement.

A study of Table 3 suggests that no significant events took place which were not identified by the regression formulas, i.e., the rate case was flagged, and the models identified unusual items while not requiring investigation of outliers explained by fluctuations in production, plant, expansion, and similar operating characteristics. While the traditional review technique typically requires investigation of all net account changes beyond 5%, the regression approach requires investigation only if known "interrelationships of elements of financial information" do not account for the recorded book values. Hence, regression conforms better to the audit program directives as to the desired approach to analytical review than does traditional variation analysis.

ANALYTICAL INVESTIGATION OF RESIDUALS

The focus in Table 3 was to investigate all outliers beyond the bounds of a 95% prediction interval. Yet, this decision criterion is relatively arbitrary, simply implying that only 5% of all outliers investigated will be attributable to random chance. In fact, the use of a 90% prediction interval would be more conservative from an auditor's perspective, since more outliers would be "checked out"; however, investigation costs would climb since 10% of the outliers investigated would be expected to stem from random
fluctuations. While the loss function of the auditor (see Wallace, 1979) supports a conservative approach to investigating outliers, the cost of additional sampling and investigation must be balanced against the cost of not finding a misstatement. This issue is left for future research.

After investigating outliers, based on some prediction interval criterion, the question arises as to what size of a misstatement can be expected for book values which are not outliers? The rule of thumb provided to the auditor is that the possible misstatement at a 95% confidence level, is no less than the model precision for the base period regression estimates. The exact boundary of the possible misstatement (as opposed to only a floor) is reported as 1/2 the 95% prediction interval per regression prediction for the audit period. This value is the auditor precision actually obtained which, in turn, can be compared with the auditor's desired precision (1/2 materiality, in units of measure identical to the variable being explained by the regression model). By comparing the 1/2 prediction interval, observed outliers, and desired precision, the auditor can assess those outliers which have deviations within the desired precision range and those estimates which, although not identified as outliers, do not have the desired resolving power, since their possible error exceeds that which would be desirable for the review or audit engagement. Additional investigation or alternative tests would be required to compensate for the excess of possible error over desired error at a 95% (or other specified) level of confidence.

An approach to residual investigation, largely devoid of outlier definitions in terms of prediction intervals, is the equiprobable residual technique described by Kinney (1979). Essentially, standardized residuals
direct the order in which residual excesses are investigated, until a sum equal to the difference in actual and desired precision is included. Although Kinney qualifies the approach—by stating the need to evaluate the relative cost of investigation versus extended tests of details, as well as recognizing that the difference in desired and actual precision can stem from a large error, a large standard error, or some combination of the two—the approach fails to explicitly recognize the inherent limitations of high standard error regression models.

Kinney (1979) also suggests a combination of the prediction interval approach and the equiprobable residual investigation technique. A filter on individually large residuals is included in the investigation process, since large residuals are considered to be indicators of situations of potential audit significance.

Although Kinney's approach was directed at an annual audit, with precision expressed in annual terms, and the standard error of the model annualized for comparison, it can be adapted to varying time frames. Table 4 reports the method of computation and the residual quantities which would be subjected to further investigation under each of the audit approaches described herein for the first quarter's limited review. The table bears out the importance of Kinney's warning that

the auditor must assess the probability of obtaining a $\bar{\sigma}$ (standard error) low enough to allow a sufficient reduction in tests of details sample size to justify the expected cost of developing and testing a regression model. (1979, p. 474)
### Alternative Approaches to Residual Investigation

<table>
<thead>
<tr>
<th>Approach 1</th>
<th>Approach 2</th>
<th>Approach 3: Equi-probable Residual Approach (see Kieney, 1979)</th>
<th>Approach 4*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Auditor-Specified Model:</strong></td>
<td><strong>1/2 (95% Prediction Interval) or Auditor Precision Less Desired Precision</strong></td>
<td><strong>Residual/Std. Error</strong></td>
<td><strong>Standardized Residual</strong></td>
</tr>
<tr>
<td><strong>Selected for Base Period (pre-Table 3 Adj.)</strong></td>
<td><strong>E [Outlier]</strong></td>
<td><strong>Distribution of Total Excess</strong></td>
<td><strong>Distribution of Total Excess</strong></td>
</tr>
<tr>
<td><strong>Electric Revenue Model</strong></td>
<td><strong>u = 5460.95</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J. Below 672.41</td>
<td>3901.99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F. Below 1108.03</td>
<td>5956.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M. Above 8809.49</td>
<td>8524.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E = Above 7029.05</td>
<td>11972.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Interest--LTD</strong></td>
<td><strong>u = 361.90</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J. Below 57.04</td>
<td>285.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F. Below 290.81</td>
<td>371.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M. Below 1101.7</td>
<td>429.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E = Below 1449.6</td>
<td>1185.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Gas Revenue</strong></td>
<td><strong>E = 1779.54</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J. --</td>
<td>1783.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F. --</td>
<td>1680.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M. --</td>
<td>1874.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E = --</td>
<td>1588.63</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Production Expense--Electric</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J. --</td>
<td>4290.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F. --</td>
<td>4505.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M. --</td>
<td>4299.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E = --</td>
<td>6254.18</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Investigation Within Prediction Intervals**

Suggested by Comparing Auditor’s Model Precision with Desired Precision

| **Gas Revenue** | 1783.72 | 533.72 | 138.57/883.0 | -.15 |
| **Production Expense--Electric** | 4290.06 | 2058.06 | 1168.46/2122.8 | .55 |
| | 4505.85 | 2256.83 | 1873.23/2231.1 | .84 |
| | 4299.29 | 1959.29 | 3151.06/2083.8 | 1.51 |

504.7 Excess Not Allocated to Residuals

To Be Investigated

| **J:** | 133.15 | **F:** | 231.63 | **M:** | 719.15 |
| **Production Expense--Electric** | 4290.06 | 2058.06 | 1168.46/2122.8 | .55 |
| | 4505.85 | 2256.83 | 1873.23/2231.1 | .84 |
| | 4299.29 | 1959.29 | 3151.06/2083.8 | 1.51 |

61.51 Excess Not Allocated to Residuals

To Be Investigated
### Table 4 (continued)

<table>
<thead>
<tr>
<th>Approach 1</th>
<th>Approach 2</th>
<th>Approach 3: Equiprobable Residual Approach (see Kinney, 1979)</th>
<th>Approach 4*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Auditor-Specified Model:</strong></td>
<td><strong>Outliers Beyond Bounds of 95% Prediction Interval</strong></td>
<td><strong>Residual/Std. Error</strong></td>
<td><strong>Distribution of Total Excess</strong></td>
</tr>
<tr>
<td><strong>Selected for Base Period (pre-Table 3 Adj.)</strong></td>
<td><strong>1/2 (95%) Prediction Interval</strong></td>
<td><strong>and Auditor Precision</strong></td>
<td><strong>Standardized Residual</strong></td>
</tr>
<tr>
<td></td>
<td>F. --</td>
<td>1216.52</td>
<td>416.52</td>
</tr>
<tr>
<td></td>
<td>M. --</td>
<td>1357.86</td>
<td>557.86</td>
</tr>
<tr>
<td>E = --</td>
<td>u = 1241.37</td>
<td>E = 1324.11</td>
<td></td>
</tr>
<tr>
<td>To Be Investigated</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allowance for Borrowed Funds</td>
<td>J. --</td>
<td>505.9</td>
<td>305.9</td>
</tr>
<tr>
<td>Used During Construction (APDC)</td>
<td>F. --</td>
<td>548.9</td>
<td>348.9</td>
</tr>
<tr>
<td></td>
<td>M. --</td>
<td>824.3</td>
<td>624.3</td>
</tr>
<tr>
<td>E = --</td>
<td>u = 624.4</td>
<td>E = 1279.1</td>
<td></td>
</tr>
<tr>
<td>46.3 Excess Not Allocated to Residuals</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>To Be Investigated</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| SUMMARY: |

Residual Investigation Approaches Include:

- Investigate all outliers beyond the bounds of the 95% prediction interval (Approach 1)
- Investigate outliers beyond the bounds of the 95% prediction interval if the sum of 1/2 the interval and the outlier exceed the auditor's desired precision (Approach 2)
- Investigate equiprobable residuals until either the desired precision is reached or no residuals remain to be explained (Approach 3)
- Investigate according to Approach 3 plus investigate all outliers beyond the bounds of the 95% prediction interval (see Kinney, 1979--addition of filter) (Approach 4)

<table>
<thead>
<tr>
<th>Outliers Investigated:</th>
<th>Approach 1</th>
<th>Approach 2</th>
<th>Approach 3</th>
<th>Approach 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Electric Revenue</td>
<td>-672</td>
<td>-1108</td>
<td>--</td>
<td>64.25</td>
</tr>
<tr>
<td>Gas Revenue</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Production Expense--Electric</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Production Expense--Gas</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>APDC</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Interest--LTD</td>
<td>-57</td>
<td>-291</td>
<td>-1102</td>
<td>--</td>
</tr>
</tbody>
</table>

*Combine Approach 1 and Approach 3
For the equiprobable residual approach, the use of a tight desired precision results in the investigation of almost all residuals, as well as an excess that is unexplained by observed residuals. Rather than checking the standard error for the base period and predictions to assess the practicable resolving power of the regression model under study, the technique prescribed in existing literature presents no cut-off rules on when to stop investigating standardized residuals due to inherent limitations of the auditor-specified base model in generating narrow prediction intervals. Kinney (1979) refers to a potential ceiling of 3 out of 12 excesses being subjected to residual analysis on a cost-beneficial basis before deciding, instead, to perform external tests of details; this is the sole rule of thumb provided—clearly of little use in the quarterly limited review setting.

Further research is required as to the preferred approach to residual investigation. Even when a regression model does not achieve the desired precision, the auditor can make use of the regression outliers as indicators of trouble-spots and as a basis for time stratifying, segmenting, or partitioning (see Kinney, 1979, p. 466) the audit population. However, the optimum approach to utilizing regression models with varying resolving power will require numerous field experiments. This study serves only to highlight the requirement that the base model precision and future prediction intervals be formally integrated into an auditor residual investigation procedure.

**STEPWISE MODEL SELECTION**

Given the primary role of standard error in determining the residual investigation approach of auditors, an attempt was made to minimize standard
error via a stepwise (optimum) regression (see Maddala, 1977, p. 126) program for model selection. However, this attempt is not intended to recommend a stepwise variable selection technique as the primary means of specifying regression models for the audit setting. Stepwise selection procedures are frequently misused in empirical work as mechanical devices that let the computer pick up the variables it likes, with no clear objective. An equation with a higher $R^2$ is not necessarily better than an equation with a lower $R^2$; at each stage of variable selection the coefficients should be examined to determine whether they have the right signs and whether they make sense and the residuals should be analyzed as well. Since the stepwise procedures invalidate the usual statistical tests of significance due to repeated tests on the same data as a basis for model selection (Maddala, 1977, p. 127), there is added difficulty in interpreting tests of the final model and constructed confidence intervals. While stepwise procedures can be useful when applied properly, it appears the auditor's objectives imply that a hypothesized model should be taken as a "given" drawn from knowledge of a client's operations and past audit experience, with statistical measures used to support the variable selections, rather than to search for the descriptor variables.

The sole application of a search technique which appears to be reasonable would be selection between substitute measures. However, the auditor should be concerned, even then, with testing whether the model changes substantially across the substitute measures under consideration and explaining why such changes are observed (particularly since high multicollinearity can be expected to affect the stepwise variable selection process). Frequently, the
alternative approach of selecting between substitutes by first testing the measures' correlation to assure the substitute qualities of the variables under consideration and then by selecting that measure most likely to have less measurement error is preferred to a stepwise procedure.

The drawbacks of applying stepwise procedures as a primary model-building device stem not only from concerns over the $R^2$ criterion, but also from the auditor's intended ongoing application of developed models which requires a well-specified theory-based relationship. If a stepwise procedure were applied annually and different variables were selected for the model, it would not be sufficient from an auditing perspective to simply "explain" that the explanatory power is higher for the new model.

- Would the variables determined to be significant last year still be significant this year, if not replaced by a new substitute variable entered by the auditor?
- If a productivity measure is no longer significant, does this imply misrepresentation of actual operations by accounting numbers?
- If a segment of business is no longer of significance in describing total revenue, are there firm assets being extracted or has the management intentionally shifted its lines of business?

The point is that changes in a model across time are as relevant as the difference observed in predicted and actual book values for the period under examination. While stepwise can be supported as a model-building aid in the first year's audit of a client, in subsequent engagements, based on in-depth knowledge of client operations, the auditor should specify the regression relationship and explain the justification for any model changes through time. Subsequent years' stepwise applications should be carefully monitored to prevent abuse, and most likely are primarily useful in checking the
reasonableness of the standard error achieved via auditor-specified models. Analysis of model changes will be performed in subsequent field applications for the utility client under study; however, the relative performance of auditor-specified and stepwise-selected models for the first quarter review engagement will now be considered.

Table 5 reports those variables subjected to stepwise, the descriptor variables selected, and the relative precision and outlier findings of the stepwise and auditor-specified regression models. The stepwise procedure was applied to the 84-observation levels form of the regression models to identify the key variables; then those descriptor variables were estimated in the same model form (i.e., number of observations and the level vs. first-differenced variable form) as that selected for the auditor-specified models to ensure comparability of standard errors. The base model precisions for the stepwise-selected models, as reported in Table 5, compare favorably to the desired precision for the electric revenue, gas revenue, gas production expenses, depreciation, and interest--LTD accounts. Similarly, the stepwise results improve all of the auditor-specified models, with respect to standard error, except for electric production expenses and AFDC. However, it is well recognized that stepwise procedures can "overfit" data. Further, the litigation environment of the auditor guarantees that occasions will arise in which the statistical models applied by an auditor will undergo careful scrutiny. An explanation of why specific variables were selected and the rationale for the direction of observed coefficients will be required. The absence of certain variables from the models will be questioned, as will changes in the formulated models across time. Deficiencies in model development
<table>
<thead>
<tr>
<th>Variables Subjected to Stepwise</th>
<th>Selected Model Form</th>
<th>Model Precision: @ 95% (Desired Precision) [Precision for Auditor-Specified Model]</th>
<th>Outliers (Differences from outliers identified by Auditor-Specified Model)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gas Revenue = f(Number of Customers--Gas Residential, * Number of Customers--Gas Other, Degree Days, * CPI, * Production Expense--Gas, * Budgeted Gas Revenue, * Gas and Electric CPI Distribution Expense--Gas, Gas Production Expense (other), Net Income, Gas Plant, Common Plant, Accounts Receivable, Accounts Payable, Customer Deposits, MCP's--Residential, * MCP's--Other, Season 8 Dummy)</td>
<td>1.78 (13.3) MCP's--Res.* .66 (1.3) Cust.--Gas Other -5.28 (-7.7) Degree Days* 1.10 (17.7) Prod.Exp.--Gas* -.04 (-1.1) Budg.Gas Rev.* .03 (2.0) Gas Prod.Exp. (other) .11 (2.9) Net Income -17452.8 (-1.4) Constant</td>
<td>878.6 Base (1250.) [1068.] 1521.1 J. 1594.7 F. 1582.6 M.</td>
<td>None (No Difference)</td>
</tr>
<tr>
<td>Gas Production Expense = f(CPI, * Gas Revenue, * Net Income, * EPS-CS, * MCP's--Residential, * Degree Days, * Gas and Electric CPI, Budgeted Production Expense--Gas, Distribution Expense--Gas, MCP's--Other, Gas Production Expense (other), Gas Plant, Common Plant, Accounts Receivable, Accounts Payable, Customer Deposits, Number of Customers--Gas Residential, Number of Customers--Gas Other, Season 1 Dummy)</td>
<td>.83 (39.3) Gas Rev.* -.11 (-3.7) Net Income* -1.3 (-15.0) MCP's--Res.* 5.7 (11.8) Degree Days* -645.6 (-2.9) Constant</td>
<td>765.18 Base (800.) [782.] 1044.3 J. 1087.3 F. 1211.9 M.</td>
<td>None (No Difference)</td>
</tr>
<tr>
<td>Variables Subjected to Stepwise</td>
<td>Selected Model Form Coefficient (t-value)</td>
<td>Model Precision # 95% (Desired Precision for Auditor-Specified Model) Monthly 1/2 (95% Pred. Interval)</td>
<td>Outliers (Differences from outliers identified by Auditor-Specified Model)</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------</td>
</tr>
<tr>
<td>AFDC = f(Budgeted AFDC, CPI, Construction W-I-P, Interest LTD, Long-term Debt--Mortgage Bonds, Dividends--P.S., Utility P.S. Yields, Utility Bond Rates, Electric Plant, Gas Plant, Common Plant, Construction Funds Held in Escrow, Sinking Fund Requirement, MEF's--Gas Residential, MEF's--Gas Other, KWH's--Residential, KWH's--Other, Production Expense--Electric, Production Expense--Gas)</td>
<td>.00 (2.4) KWH's--Res. .00 (.0) Prod.Exp.--Elec. .13 (2.0) Budge.AFDC* .00 (2.4) Constr.W-I-P* -.59 (-1.5) Div.--P.S.* 979.5 (1.2) Constant</td>
<td>135.34 Base (200.1) [332.1] 418.4 J. 421.8 F. 465.9 M.</td>
<td>None (No Difference)</td>
</tr>
<tr>
<td>Interest--LTD = f(CPI, Budgeted Interest LTD, AFDC, Net Income, Dividends--P.S., Utility Bond Rates, Utility P.S. Yields, Interest S-T Debt, Net Amortization of Debt Premium Discount &amp; Expense, Electric Plant, Construction W-I-P, Unamortized Debt Discount &amp; Expense, Long-Term Debt--Mortgage Bonds, Amortized Premium on LTD, Unamortized Discount on LTD, Sinking Fund Requirement, Interest Accrued, Linear Trend)</td>
<td>.77 (10.9) Budge.Int.--LTD* .07 (.3) AFDC* .01 (4.7) Net Amort. of Debt Prem.Disc. &amp; Exp. .05 (3.9) Sinking Fund Requirement 1141.5 (2.4) Constant</td>
<td>150.8 Base (450.0) Runs Sign. # 95: Runs Ratio -1 [212.9] 172.0 J. 170.9 F. 179.5 M.</td>
<td>M. Below 717.2 (J. Below F. Below Also Identified)</td>
</tr>
</tbody>
</table>

*Variable Included in Auditor-Selected Model
which could have been avoided with "due audit care" and inadequate follow-up of model changes could greatly impair an auditor's legal defense. The implications of the litigation environment and the auditor's objectives of understanding as opposed to simply forecasting a client's operations suggest the impropriety of the auditor applying regression-selection procedures such as stepwise regression as the primary means of model development, despite the lower standard error available via such a technique. This lower error is primarily useful as a benchmark for assessing the performance of auditor-specified models.

The model precision for the auditor-specified relationships is looser than for the stepwise models in five of the seven regressions: by 2%, 14%, 18%, 22%, and 29%. Since the variables subjected to the stepwise procedure were reasonably expected to have some relationship to the accounts being analyzed, no surprising relationships result in the selected models. Given the limited training of the audit manager and the absence of field experience in applying regression, it can be reasonably expected that several of these variables may have suggested themselves to the auditor as additional key determinants of fluctuations in the variables of interest, had he the benefit of greater model-building expertise. Considering the absence of an eighteen-variable search process (on average), and the substantial overlap of selected descriptor variables for the auditor-specified and stepwise-selected models, this field experience supports the auditor's ability to identify key regression relationships without the assistance of stepwise.

Further, there are no outliers identified by stepwise which are not also identified by the auditor-specified model, although four additional outliers are apparently explained by the stepwise regression formulations.
Returning to the adequacy of the auditor-specified model, when each model's precision per prediction is compared with the desired auditor precision, eight out of twenty-one values reach the desired resolving power to justify no further audit work. The remaining values suggest some detailed testing would be required to augment the regression findings. However, the audit work required to lower the 1521 January value of possible error in gas revenue, for example, to the desired 1250 level, can be expected to be less than the testing required without the partial resolution power of regression.

COST/BENEFIT FACTORS AND 
REGRESSION'S EFFECTS ON AUDIT PLANNING

As already suggested by the comparison of traditional variation analysis and regression as tools for limited review, the regression approach substantially decreases the amount of investigation required by the auditor. When the auditor achieves the desired precision, (s)he can limit investigation to outliers. This means that, in many cases, the explanation for a 15% increase in sales, or some similar account fluctuation, will be that it is fully explained by the regression model. From a working paper preparation point of view, this means that the auditor does not have to write a lengthy explanation (which is frequently redundant across the separate traditional variation analyses) for a situation which is normal.

1. Cost/Benefit Factors

However, in assessing the effects of regression on both the limited review and subsequent audit planning process, it is essential that direct costs of regression, including computer and training costs be quantified. Regression requires data bases to be constructed in machine-readable form.
Due to relatively cheap data entry mode charges, a typical file could be created in at most one hour per variable, meaning a computer charge of $300 for 20 variables. In most cases, input can be done by a secretary. The actual time-sharing application of regression costs roughly $2.80 per run. Obviously, the computer costs are nominal.

The biggest cost relates to training. It is difficult to measure the cost of training and differences in training requirements between traditional analytical review and regression analysis for the field experiment. This is because there is currently no formal Price Waterhouse & Co. (PW) training program in analytical review techniques. Instead, staff members are taught to perform a fluctuation analysis (to compare amounts between periods and attempt to find reasons for the changes). They also learn about analytical review through on-the-job training.

The training requirements to apply regression analysis include the following. First, the auditor needs to have some understanding of statistical sampling techniques and the use of timesharing. As a practical matter, this means that (s)he must be a graduate of the 3-day introductory seminar provided by PW which teaches how to use attributes and variables sampling and familiarizes the auditor with the use of time-sharing for statistical sampling. On top of that, (s)he needs at least a 2- to 3-day training program in analytical review using regression analysis. Some of this training will undoubtedly be necessary just to teach the auditor the fundamentals of analytical review whether using regression analysis or nonstatistical methods. (This is similar to courses on statistical sampling, where the auditor learns more about auditing than statistics.) The major training cost is in the development of courses and the loss of billings when the staff attends training.
With such computer and training cost factors identified, the other
direct cost involves the total audit time required for a regression appli-
cation. From two field experiences at PW & Co., it is estimated that 40
to 50 hours of staff time and 10 to 20 hours of manager time are required
for a regression application. This estimate would vary with the number of
models constructed and the familiarity of the audit team with the client.
In order to specify useful audit relationships, the auditor must understand
a good deal about the business. (S)He must understand the client's business
cycle, that is how long a period does it take to go from cash to inventory
to receivables and back to cash. Such an understanding permits the appro-
priate construction of lagged variables in the regression models. The
auditor must know where, when, and how to obtain and use variables from
outside the business and when internal data is more appropriate. The de-
sire for external data means the auditor must get out of the accounting
department, into the marketing, forecasting, and production departments.
(S)He must also become aware of how the client creates its budgets and
whether the budgeting practice is effective.

The type of information needed to properly apply regression analysis
is normally known only by the manager and partner. In many cases, it may
not be known to the staff accountants. Thus, to properly apply regression
analysis, there must be active participation by the manager and partner.

2. Effects on Audit Planning

The balance of this paper is directed to a limited review application
of regression and its effect on the quantity of variations analyzed. At
some point, such an application will be integrated into the audit planning
process and presumably affect the sample sizes selected for substantive
testing. However, before one can use regression in audit planning, one must
accept the premise that an analytical review is considered a form of substantive testing and that it provides valid audit evidence. It is not just a final check at the end of the audit but is an acceptable alternative to compliance and substantive testing. (There are still some auditors who believe that analytical review is not audit evidence.)

Once it is agreed that analytical review constitutes a valid form of substantive testing, the regression procedure can be integrated with other compliance and substantive tests. Figure 3 presents an overview of how to evaluate the role of regression models constructed during a limited review (or additional models) in the audit process. The figure focuses on the central criterion of model precision and integrates results and examples from the field experiment, as well as a related analytical paper by Kinney (1979). The total effect on sample sizes of such an integration procedure is estimated to lead to a 5% decline in overall audit hours. Such savings stem from the availability of useful regression models which can lead to the partitioning of the audit population and, in some cases, no additional detailed testing would be required. The Appendix indicates 4 outliers in the second quarter out of 21 predictions (7 models, with 3 months per model). If this ratio of outliers persists, Kinney's (1979) rule of thumb that up to one-fourth of the predictions could have outliers and still be investigated in a cost/beneficial manner, relative to an extension of detailed tests, will be met. In that case, the investigation could suffice to meet the desired auditor precision for several of the models, or, at least, lower the estimated possible misstatement of the accounts being reviewed. If, instead of an investigation procedure, detailed tests were performed which reflected the regression findings, and if Kinney's (1979) dollar unit sampling simulation
*Kinney (1979) reports a decrease in Dollar Unit Sample Size of 59% when regression is used with partitioning and 33% without partitioning.

Is P.I. ≤ D.P.? NO

*Note that a<br>reliable<br>approach<br>can<br>result in a<br>big chance of<br>over-auditing<br>errors caused<br>by random<br>fluctuations<br>due to<br>large standard<br>error in the<br>regression<br>base model.

Is the E[PI and observed outliers] ≤ D.P.? NO

*For the utility client, specific applications of regression analysis to select which month's deferred fuel costs to recompute, and to provide a basis for partitioning monthly for detailed testing of material issues, disbursements, and construction work-in-process would decrease audit risk exposure in these account areas.

Estimate regression models and either investigate outliers or partition population for detailed testing.

Estimate regression models and utilize for planning purposes.

Estimate regression models and compare the estimates to the audited statements.

*This suggests a possible use of regression for limited review (with looser precision requirements), despite limitations on regression's use for the audit engagement.

DO NOT APPLY THE REGRESSION TOOL FOR THE MODEL(S) CONSTRUCTED IN THE LIMITED REVIEW. HOWEVER, CONSIDER ALTERNATIVE APPLICATIONS AND EVALUATE THEIR AUDIT USE THROUGH THE SAME DECISION PROCESS.

*The auditor should recognize that regression need not use every variable or be used for every account. Cut-off problems for some data may not relate to other accounts. Thus, a regression might be useful to test payroll but not purchases. For the utility client, regression applications for Pension (ERISA) analysis and payroll tests are likely to be effective.

Investigate outliers or, partition population (Kinney, 1979) and perform detailed tests.

Auditor's judgment is required to assess whether large outliers should be investigated despite their values ≤ D.P.
results are to any degree generalizable to 2-sided alternative sampling approaches, drops in sample size approaching 60% can be anticipated.

The field experiment leaves numerous unanswered questions as to how the regression application for limited review will affect the audit and selected sample sizes. This evidence must await completion of the annual audit and further analysis. However, the indications provided by critical analysis of past auditing procedures, existing literature, available regression model precision, and the frequency of outliers over the first two quarters, are that substantial savings in audit time and efficiency in sample selection can result from applying the regression procedure.

LIMITATIONS AND CRITICAL AREAS FOR FUTURE RESEARCH

Since the extent of problems undetected by either traditional analytical review procedures or regression analysis is unknown, this field experiment provides only a comparison of the two techniques with respect to known errors and observable effects on audit planning. Further, as with any case study, the ability to generalize the results is limited. However, while only one company was subjected to a review, seven account balances were analyzed and 84 historical monthly data points were available for model construction. This suggests that there was an adequate database for effective model-building from a statistical perspective and that consistency in model-building inferences across accounts can increase the strength of inferences drawn from the limited review application.
Numerous research questions for future study are suggested by this field experiment:

- How generalizable are results concerning the frequency of statistical problems and capability of transformations to "fix-up" base models.
- The relative performance of regression analysis and traditional review techniques.
- The relative performance of stepwise and auditor-specified models.
- The observed magnitudes of model precision achieved by auditor specification of regression relationships.
- The frequency of outliers.
- What trade-off's in model building precision and quality of evidence accrue from varying models with respect to the inclusion of external data, budgeted data, and other internal accounting data.
- What is the cost/benefit optimum way to approach data collection with respect to the number of observations included in a base model.
- What power does Cochrane-Orcutt have in improving model precision in the presence of autocorrelated residuals.
- What method of residual evaluation is optimal.
- How can the standard error limitations of constructed models be integrated with the Kinney (1979) approach.
- How important is a filter that focuses on large residuals, or perhaps outliers with respect to a set prediction interval.
- How should the 1/2 prediction interval direct auditors' reliance on regression results.
- Given the evidence as to model precision and the frequency of outliers.
- What advantages in terms of sample size can be expected to accrue from the use of regression, particularly the use of regression for partitioning an audit population.
- How does the resolving power of regression affect the overall planning process.
- How frequently do different models signal different outliers with no apparent explanation and what are the related audit implications.

Most of these research questions require extensive replications of the reported field experiment and simulation studies. However, to understand the acceptability of regression findings and their integration with more
traditional audit techniques, behavioral studies on the audit planning process and the varied perceptions of what constitutes audit evidence and what resolving power of regression is essential for its use in the audit and/or limited review engagement could be useful. A practical problem with any new audit technique is that auditors tend to be extremely conservative and somewhat reluctant to try anything new. As a practical approach to assessing the role and potential of regression as an audit tool, substantial research is required to demonstrate its reliability, limitations, and cost/benefit contribution to the auditing process.

Once the role and potential of regression is established, a second practical problem of where and when to apply regression arises. Regression analysis is not appropriate for all clients. For example, regression analysis should not be used in a time series manner at a client that has had major changes either in the recent base period or the audit period. Also, if a client operates in many lines of business, the availability of data on a segmental basis will determine how difficult it is to apply regression. Generally, time series regression will work best for clients that are relatively stable, or at least, in audit areas which are relatively stable for a given client. For such clients the performance of regression will rest, in part, on the measurement error of the data base. Most economic data contain observational errors and the variables the auditor actually measures are imperfect measurements of what (s)he wants to measure. The error may stem from poor cut-offs. If the auditor knows the source of errors, it may be possible to adjust the data (for example, by adjusting the cut-off or by correcting for "passed" or "unbooked" adjusting journal entries (AJE's) proposed by the auditors during their examinations in prior
periods) and thereby minimize problems with measurement error and its
potential distortion of the regression model. However, if the measurement
error is deemed to be severe and unadjustable, such data cannot be used
as regression variables. Similarly, a careful evaluation of internal con-
trol and subsequent model specification that reflects endogeneity concerns
will facilitate the selection of appropriate variables. Future field
applications should provide a basis for specifying a list of factors to
consider in determining whether the client's operations will permit an
effective regression application. Often when instability imperils the
potential of time-series applications, cross-sectional applications will
still be useful to the auditor as a means of selecting clients' operating units
for testing. As depicted in Figure 3 regression can be applied to address
various audit questions, depending on limitations of a specified model as well
as reservations an auditor might have in relying on the regression analysis--
due to measurement error, endogeneity, data availability or similar model-
building concerns. The timing of a regression application will largely depend
on the audit question being addressed. Again, additional research is required
to direct the application of regression to the audit.

**SUMMARY**

Actual field experience with regression analysis supports its effect-
iveness in a limited review setting with respect to identifying similar
fluctuations to traditional variation analysis and adding a historical
perspective to such an analysis that provides insights regarding shifts in
management policy and underlying operations. Statistical issues in model-
building, regression software design specifications, trade-offs in model
precision that result from the use of different types of data and varied approaches to formulating models, and a comparison of stepwise regressions to auditor-specified models are presented. The regression tool is utilized to check the reasonableness and sufficiency of explanations for fluctuations noted in the traditional variation analysis; when the expansion of the regression models do not fully explain observed fluctuations, additional investigation is performed and potential problem areas (i.e., a shift in financing policy) are detected which were not identified in the traditional review process.

When the limited review results are analyzed by the audit team in determining audit scope, the availability of regression base models with desired precision levels and a moderate number of outliers is considered a basis for reducing detailed tests in 6 of the 7 accounts reviewed and for decreasing the required audit time for both subsequent limited review engagements and the revenue/expense analysis performed as part of the annual audit. The total level of required investigation of fluctuations is observed to be higher for traditional variation analysis, since the investigation of substantial percentage shifts by account ignores well known operating fluctuations, whereas regression explicitly controls for such factors, much as flexible budgeting adapts to operations in a cost accounting system. Only an overall estimate of audit time savings is provided, since the audit planning for the client under study was at a preliminary stage at completion of the field experiment. Studies are currently in process which will more directly consider audit planning issues, and will apply the regression technique to other field settings to derive some measure of the ability to generalize research results reported herein.
APPENDIX

Footnotes for Tables A through G

*95%
**99%
I = Inconclusive @ 95% with respect to positive autocorrelation
II = Inconclusive @ 95% with respect to negative autocorrelation
N = Negative autocorrelation @ 95%
P = Positive autocorrelation @ 95%
F = A 91% value, Table E; 89.9%, Table F
<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Coefficient (Standard Error)</th>
<th>t-Statistic</th>
<th>Models</th>
<th>Levels 36 obs.</th>
<th>First Differences 35 obs.</th>
<th>Levels 84 obs. SELECTED</th>
<th>First Differences 83 obs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-672.07 (9270.98)</td>
<td>-0.732</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kWh-Residential</td>
<td>0.0108 (0.011)</td>
<td>.992</td>
<td>.0321</td>
<td>(.0147)</td>
<td>2.1832*</td>
<td>.0762</td>
<td>(.0064)</td>
</tr>
<tr>
<td>Degree Days</td>
<td>2.898 (1.6038)</td>
<td>1.711</td>
<td>.5616</td>
<td>(2.443)</td>
<td>.2299</td>
<td>.026</td>
<td>4.089**</td>
</tr>
<tr>
<td>Budgeted Electric Revenue</td>
<td>0.1725 (1.1344)</td>
<td>1.2835</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CFI</td>
<td>188.9087 (42.134)</td>
<td>4.343**</td>
<td>599.46</td>
<td>(733.54)</td>
<td>2.8172</td>
<td>161.51</td>
<td>(32.412)</td>
</tr>
<tr>
<td>Production Expense--Electric</td>
<td>.851 (.9904)</td>
<td>5.415**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model Precision at 95% Confidence [i.e., Standard Error x (1.96)]</td>
<td></td>
<td></td>
<td>3920.56</td>
<td>4927.90</td>
<td>3582.71</td>
<td>4395.19</td>
<td></td>
</tr>
<tr>
<td>Auditor's Desired Precision</td>
<td>5000.</td>
<td>5000.</td>
<td></td>
<td></td>
<td>5000.</td>
<td>5000.</td>
<td></td>
</tr>
<tr>
<td>( R^2 )</td>
<td>.9656</td>
<td>.691</td>
<td></td>
<td></td>
<td>.9992</td>
<td>.692</td>
<td>(.672)</td>
</tr>
<tr>
<td>[F-ratio]</td>
<td>(.9584)</td>
<td>(.638)</td>
<td></td>
<td></td>
<td>(.9890)</td>
<td>(.672)</td>
<td></td>
</tr>
<tr>
<td>Durbin-Watson</td>
<td>1.896</td>
<td>2.6676**</td>
<td></td>
<td></td>
<td>1.6707\textsuperscript{1}</td>
<td>2.8273\textsuperscript{1}</td>
<td></td>
</tr>
<tr>
<td>Runs Test</td>
<td>Not Significant</td>
<td></td>
<td></td>
<td></td>
<td>Not Significant</td>
<td></td>
<td>Not Significant</td>
</tr>
<tr>
<td>Obs.</td>
<td>17.</td>
<td>26.</td>
<td></td>
<td></td>
<td>45.</td>
<td>56.</td>
<td>3.</td>
</tr>
<tr>
<td>Exp.</td>
<td>18.9</td>
<td>19.</td>
<td></td>
<td></td>
<td>42.98</td>
<td>43.</td>
<td></td>
</tr>
<tr>
<td>Runs Ratio</td>
<td>--</td>
<td>22</td>
<td></td>
<td></td>
<td>--</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>Autocorrelation (lag 1) of the Residuals</td>
<td>.0167</td>
<td>-.386</td>
<td></td>
<td></td>
<td>.1541</td>
<td>-.4189</td>
<td></td>
</tr>
<tr>
<td>Autocorrelation (lag 2-12) of the Residuals in Excess of .5/#</td>
<td>-.410/4</td>
<td>.344/9</td>
<td>.451/6</td>
<td>-.530/7</td>
<td>-.399/9</td>
<td>.245/12</td>
<td></td>
</tr>
<tr>
<td>Kolmogorov-Smirnov Statistic</td>
<td>.1365</td>
<td>.0931</td>
<td></td>
<td></td>
<td>.0552</td>
<td>.0612</td>
<td></td>
</tr>
<tr>
<td>Residuals: Skewness Kurtosis</td>
<td>.4515</td>
<td>2.708</td>
<td>.2588</td>
<td>4.5914</td>
<td>.3985</td>
<td>4.0699</td>
<td>3.92</td>
</tr>
<tr>
<td>Outliers</td>
<td>None</td>
<td>None</td>
<td>Below 3729.7</td>
<td>Below 672.4</td>
<td>Below 1108.02</td>
<td>Above 8809.49</td>
<td>None</td>
</tr>
<tr>
<td>1st Quarter January</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>February</td>
<td>None</td>
<td>None</td>
<td>Below 3729.7</td>
<td>Below 672.4</td>
<td>Below 1108.02</td>
<td>Above 8809.49</td>
<td>None</td>
</tr>
<tr>
<td>March</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>2nd Quarter</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>April</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>May</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>June</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>Independent Variables:</td>
<td>Coefficient (Standard Error) t-Statistic</td>
<td>Levels 36 obs. SELECTED (Model Shift**)</td>
<td>Levels 84 obs.</td>
<td>First Differences 35 obs.</td>
<td>First Differences 83 obs.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>------------------------------------------</td>
<td>----------------------------------------</td>
<td>----------------</td>
<td>--------------------------</td>
<td>--------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-2005.45 (.4506)</td>
<td>127.69 (334.22)</td>
<td>-7804.44 (17597.6)</td>
<td>.4453</td>
<td>155.401 (243.571)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of Gas Customers</td>
<td>.0445 (.0527)</td>
<td>.3859 (.2135)</td>
<td>.0189 (.0455)</td>
<td>.2615</td>
<td>-0.0119 (.1312)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCF's--Residential</td>
<td>2.0911 (.1423)</td>
<td>2.006 (.254)</td>
<td>1.518 (.1220)</td>
<td>1.691 (.1708)</td>
<td>9.8977**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Degree Days</td>
<td>-6.025 (.7827)</td>
<td>-5.329 (1.005)</td>
<td>-6.566 (1.7035)</td>
<td>-3.962**</td>
<td>-4.065**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPI</td>
<td>25.074 (11.603)</td>
<td>-51.670 (203.12)</td>
<td>23.928 (7.4765)</td>
<td>-51.968 (178.531)</td>
<td>-1.179</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Production Expense--Gas</td>
<td>1.129 (.214)</td>
<td>1.0571 (.1212)</td>
<td>1.0739 (.0702)</td>
<td>.6978 (.0849)</td>
<td>8.211**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Budgeted Gas Revenue</td>
<td>-0.926 (.0467)</td>
<td>-0.0515 (.0638)</td>
<td>-0.704 (.0452)</td>
<td>-1.173 (.0536)</td>
<td>3.3076**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model Precision at 95% Confidence [i.e., Standard Error x (1.96)]</td>
<td>1067.52</td>
<td>1300.68</td>
<td>1697.8</td>
<td>1840.88</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Auditor's Desired Precision</td>
<td>1250.0</td>
<td>1250.00</td>
<td>1250.00</td>
<td>1250.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$^2$</td>
<td></td>
<td>.998</td>
<td>.9973</td>
<td>.967</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$^2$</td>
<td></td>
<td>(.999)</td>
<td>(.997)</td>
<td>(.964)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Durbin-Watson</td>
<td>[2179.5]**</td>
<td>[462.39]**</td>
<td>[1750.7]**</td>
<td>[371.18]**</td>
<td>2.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Runs Test</td>
<td>Not Significant</td>
<td>Not Significant</td>
<td>25.</td>
<td>33.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obs. #</td>
<td>19.</td>
<td>20.</td>
<td>43.</td>
<td>43.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exp. #</td>
<td>15.</td>
<td>15.</td>
<td>43.</td>
<td>43.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Runs Ratio</td>
<td>--</td>
<td>--</td>
<td>43.</td>
<td>43.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autocorrelation (lag 1) of the Residuals</td>
<td>.1914</td>
<td>.426</td>
<td>.2426</td>
<td>-.034</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autocorrelation of the Residuals in Excess of .3/lag</td>
<td>-.327/8</td>
<td>-.479/6</td>
<td>--</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kolmogorov-Smirnov Statistic</td>
<td>.0827</td>
<td>.085</td>
<td>.1557</td>
<td>.087</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residuals: Skewness Kurtosis</td>
<td>.605</td>
<td>.267</td>
<td>3.028</td>
<td>3.167</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outliers</td>
<td>None</td>
<td>None</td>
<td>Below 103.84</td>
<td>Below 761.67</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st Quarter January</td>
<td>None</td>
<td>None</td>
<td>Below 103.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>February</td>
<td>None</td>
<td>None</td>
<td>Below 982.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>March</td>
<td>None</td>
<td>None</td>
<td>Below 53.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>-------------------------------</td>
<td>----------------------------------------</td>
<td>----------------</td>
<td>----------------</td>
<td>--------------------------</td>
<td>----------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td><strong>Constant</strong></td>
<td></td>
<td>-54,343.05 (8774.78)</td>
<td>302.15</td>
<td>(570.29)</td>
<td>-7781.65 (4120.60)</td>
<td>-1,858</td>
<td>-117,886 (312.26)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.066**</td>
<td>0.5298</td>
<td></td>
<td>0.888</td>
<td>0.3750</td>
<td></td>
</tr>
<tr>
<td><strong>Fuel CPI</strong></td>
<td></td>
<td>19.525 (51.161)</td>
<td>-281.46</td>
<td>(266.64)</td>
<td>-60.814</td>
<td>(-29.292)</td>
<td>54.106 (155.613)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.5785</td>
<td>-1,056</td>
<td></td>
<td>2.085*</td>
<td>0.3477</td>
<td></td>
</tr>
<tr>
<td><strong>Electric Revenue</strong></td>
<td></td>
<td>.7784 (.0834)</td>
<td>.7744 (.1176)</td>
<td></td>
<td>.7551 (.0603)</td>
<td></td>
<td>0.7074 (.0700)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9.3387**</td>
<td>6.5992**</td>
<td></td>
<td></td>
<td>10.099**</td>
<td></td>
</tr>
<tr>
<td><strong>Net Income</strong></td>
<td></td>
<td>-2.1570 (.7272)</td>
<td>-1.9629</td>
<td>(.8994)</td>
<td>-7.466</td>
<td>(-1.1903)</td>
<td>10.099**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0268**</td>
<td>-2.046*</td>
<td></td>
<td>-5.923**</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>EPS-C.S.</strong></td>
<td></td>
<td>320.05 (438.044)</td>
<td>246.395</td>
<td>(569.23)</td>
<td>168.701</td>
<td>(108.85)</td>
<td>1.734</td>
</tr>
<tr>
<td><strong>KWH-Residential</strong></td>
<td></td>
<td>0.0253 (.0133)</td>
<td>.0142 (.0155)</td>
<td></td>
<td>.0034 (.0073)</td>
<td></td>
<td>0.0110 (.0804)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.9066</td>
<td>2.1991*</td>
<td></td>
<td></td>
<td>1.3072</td>
<td></td>
</tr>
<tr>
<td><strong>Degree Days</strong></td>
<td></td>
<td>-5.1599 (1.5229)</td>
<td>-6.425 (2.7181)</td>
<td></td>
<td>-5.140</td>
<td>(-8.638)</td>
<td>-1.496</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-3.375**</td>
<td>-2.803**</td>
<td></td>
<td></td>
<td>1.3072</td>
<td></td>
</tr>
<tr>
<td><strong>Model Precision at 95% Confidence</strong></td>
<td>[i.e., Standard Error x (1.96)]</td>
<td>3283.01</td>
<td>4254.996</td>
<td></td>
<td>3219.44</td>
<td></td>
<td>3419.21</td>
</tr>
<tr>
<td><strong>Auditor's Desired Precision</strong></td>
<td></td>
<td>2250.</td>
<td>2250.</td>
<td></td>
<td>2250.</td>
<td></td>
<td>2250.</td>
</tr>
<tr>
<td><strong>R²</strong></td>
<td></td>
<td>.955</td>
<td>.7598</td>
<td></td>
<td>.978</td>
<td></td>
<td>.6963</td>
</tr>
<tr>
<td>(<strong>R²</strong>)</td>
<td></td>
<td>(.9466)</td>
<td>(.7084)</td>
<td></td>
<td>(.976)</td>
<td></td>
<td>(.6732)</td>
</tr>
<tr>
<td>[F-Ratio]</td>
<td></td>
<td>.9023**</td>
<td>.923**</td>
<td></td>
<td>.923**</td>
<td></td>
<td>.5950</td>
</tr>
<tr>
<td>Durbin-Watson</td>
<td></td>
<td>1.99</td>
<td>2.57211</td>
<td></td>
<td>1.2230</td>
<td></td>
<td>2.917*</td>
</tr>
<tr>
<td>Obs. #</td>
<td></td>
<td>18.78</td>
<td>18.1</td>
<td></td>
<td>43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exp. #</td>
<td></td>
<td>--</td>
<td>--</td>
<td></td>
<td>43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Runs Ratio</td>
<td></td>
<td>--</td>
<td>--</td>
<td></td>
<td>43</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Autocorrelation (lag 1) of the Residuals</strong></td>
<td></td>
<td>-.048</td>
<td>-.351</td>
<td></td>
<td>.3765</td>
<td></td>
<td>-.4754</td>
</tr>
<tr>
<td><strong>Autocorrelation of the Residuals in Excess of .3/1lag</strong></td>
<td></td>
<td>-.346/4</td>
<td>-.454/4</td>
<td></td>
<td>.4111/2</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Kolmogorov-Smirnov Statistic</strong></td>
<td></td>
<td>.1301</td>
<td>.0734</td>
<td></td>
<td>.0904</td>
<td></td>
<td>.0664</td>
</tr>
<tr>
<td><strong>Residuals: Skewness Kurtosis</strong></td>
<td></td>
<td>.435</td>
<td>.379</td>
<td></td>
<td>1.296</td>
<td></td>
<td>-3.2186</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.195</td>
<td>3.475</td>
<td></td>
<td>9.452</td>
<td></td>
<td>7.4001</td>
</tr>
<tr>
<td><strong>Outliers</strong></td>
<td></td>
<td>None</td>
<td>None</td>
<td></td>
<td>Above 796.061</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>1st Quarter</td>
<td></td>
<td>None</td>
<td>None</td>
<td></td>
<td>Above 796.061</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>January</td>
<td></td>
<td>None</td>
<td>None</td>
<td></td>
<td>Above 796.061</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>February</td>
<td></td>
<td>None</td>
<td>None</td>
<td></td>
<td>Above 796.061</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>March</td>
<td></td>
<td>None</td>
<td>None</td>
<td></td>
<td>Above 796.061</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>2nd Quarter</td>
<td></td>
<td>None</td>
<td>None</td>
<td></td>
<td>Above 796.061</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>April</td>
<td></td>
<td>None</td>
<td>None</td>
<td></td>
<td>Above 796.061</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>May</td>
<td></td>
<td>None</td>
<td>None</td>
<td></td>
<td>Above 796.061</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>June</td>
<td></td>
<td>None</td>
<td>None</td>
<td></td>
<td>Above 796.061</td>
<td>None</td>
<td>None</td>
</tr>
</tbody>
</table>
### Table D
Regression Model of Gas Production Expenses

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient (Standard Error)</td>
<td>SELECTED (Runs**=86 obs &amp; Smaller Std. Error)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t-Statistic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-516.52</td>
<td>21.637</td>
<td>-1361.09</td>
<td>49.261</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(.3223)</td>
<td>(.099)</td>
<td>(1022.65)</td>
<td>(238.89)</td>
<td>.205</td>
<td></td>
</tr>
<tr>
<td>CFI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1.885</td>
<td>2.948</td>
<td>7.384</td>
<td>-25.822</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(9.786)</td>
<td>(133.196)</td>
<td>(6.615)</td>
<td>(175.51)</td>
<td>.1271</td>
<td></td>
</tr>
<tr>
<td>Gas Revenue</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.8142**</td>
<td>.7402</td>
<td>.7547**</td>
<td>.6988**</td>
<td>9.584**</td>
<td></td>
</tr>
<tr>
<td>(.0392)</td>
<td>(.0600)</td>
<td>(.0377)</td>
<td>(.074)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Net Income</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.0422</td>
<td>.014</td>
<td>-2.778**</td>
<td>-.0008</td>
<td>.121</td>
<td></td>
</tr>
<tr>
<td>(.1942)</td>
<td>(.204)</td>
<td>(.0817)</td>
<td>(.121)</td>
<td>.501</td>
<td></td>
</tr>
<tr>
<td>EPS-C.S.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-.9578</td>
<td>-304.11</td>
<td>76.657</td>
<td>10.709</td>
<td>(60.383)</td>
<td></td>
</tr>
<tr>
<td>(120.61)</td>
<td>(115.92)</td>
<td>(47.554)</td>
<td>(130.72)</td>
<td>.177</td>
<td></td>
</tr>
<tr>
<td>MCF's--Residential</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-.1250</td>
<td>-9.107</td>
<td>-2.140**</td>
<td>-2.80**</td>
<td>.9345</td>
<td></td>
</tr>
<tr>
<td>(.366)</td>
<td>(.2304)</td>
<td>(.1311)</td>
<td>(.242)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Degree Days</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.243</td>
<td>4.854</td>
<td>6.105</td>
<td>5.179</td>
<td>7.998**</td>
<td></td>
</tr>
<tr>
<td>(.417)</td>
<td>(.591)</td>
<td>(.4077)</td>
<td>(.648)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model Precision at 95% Confidence [i.e., Standard Error x 1.96]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>782.14</td>
<td>864.73</td>
<td>1317.94</td>
<td>1788.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Auditor's Desired Precision</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>800.</td>
<td>800.</td>
<td>800.</td>
<td>800.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.988</td>
<td>.9906</td>
<td>.9905</td>
<td>.933</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(.997)</td>
<td>(.9884)</td>
<td>(.989)</td>
<td>(.927)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-Ratio</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[1027.74]**</td>
<td>[493.58]**</td>
<td>[1342.91]**</td>
<td>[174.9]**</td>
<td>2.823**</td>
<td></td>
</tr>
<tr>
<td>Durbin-Watson</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.562</td>
<td>2.4763</td>
<td>1.962</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Runs Test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obs. #</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>18</td>
<td>31</td>
<td>48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exp. #</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18.8</td>
<td></td>
<td>43</td>
<td>41.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Runs Ratio</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--</td>
<td></td>
<td>--</td>
<td>--</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autocorrelation (lag 1) of the Residuals</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.2179</td>
<td>-.255</td>
<td>.0087</td>
<td>-.424</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autocorrelation of the Residuals in Excess of .3/lag</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-.3048/9</td>
<td>--</td>
<td>--</td>
<td>.552/8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kolmogorov-Smirnov Statistic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.0879</td>
<td>.0990</td>
<td>.1353</td>
<td>.1919</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residuals: Skewness Kurtosis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.0255</td>
<td>-4.494</td>
<td>-1.723</td>
<td>-1.578</td>
<td>12.26</td>
<td></td>
</tr>
<tr>
<td>2.910</td>
<td>4.228</td>
<td>11.413</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outliers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st Quarter</td>
<td>None</td>
<td>None</td>
<td>Above 277.91</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>January</td>
<td>None</td>
<td>None</td>
<td>Above 95.9</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>February</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>March</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>2nd Quarter</td>
<td>None</td>
<td>None</td>
<td>Above 146.82</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>April</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>May</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>June</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>Independent Variables:</td>
<td>Coefficient (Standard Error)</td>
<td>t-Statistic</td>
<td>MEASURES</td>
<td>LEVELS 36 obs.</td>
<td>LEVELS 84 obs.</td>
</tr>
<tr>
<td>------------------------</td>
<td>-----------------------------</td>
<td>------------</td>
<td>----------</td>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>First Differences 35 obs. SELECTED (Kuhn's Levels &amp; Smaller Std. Error)</td>
<td>First Differences 83 obs.</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>2597.124 (1639.46)</td>
<td>1.584</td>
<td>-73.821 (43.678)</td>
<td>1.154</td>
<td>-119.75 (73.852)</td>
</tr>
<tr>
<td>Electric Revenue</td>
<td>-.00028 (.00021)</td>
<td>-.1299</td>
<td>-.0006 (.00022)</td>
<td>-.0093 (.00025)</td>
<td>-.0088 (.00057)</td>
</tr>
<tr>
<td>Gas &amp; Electric CPI</td>
<td>1.5600 (2.396)</td>
<td>.651</td>
<td>5.466 (5.051)</td>
<td>1.082</td>
<td>3.468 (4.909)</td>
</tr>
<tr>
<td>Budgeted Depreciation</td>
<td>.1391 (.0945)</td>
<td>1.466</td>
<td>.00604 (.1262)</td>
<td>.3676 (-1.036)</td>
<td>.119 (.199)</td>
</tr>
<tr>
<td>Electric Plant</td>
<td>.00022 (.00006)</td>
<td>.224</td>
<td>.0038 (.00162)</td>
<td>.0014 (.00055)</td>
<td>.001 (.0008)</td>
</tr>
<tr>
<td>Accumulated Depreciation</td>
<td>.00029 (.00010)</td>
<td>2.291**</td>
<td>.0073 (.0062)</td>
<td>-.000 (0.0012)</td>
<td>.020 (.011)</td>
</tr>
<tr>
<td>Model Precision at 95% Confidence [i.e., Standard Error * (1.96)]</td>
<td>120.251</td>
<td>143.522</td>
<td>299.98</td>
<td>384.54</td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>.952 (.944)</td>
<td>.341</td>
<td>.961 (.956)</td>
<td>.1141 (.057)</td>
<td>.1043 (.198)</td>
</tr>
<tr>
<td>[F-Ratio]</td>
<td>[118.78]**</td>
<td>1.877</td>
<td>[2.619]**</td>
<td>[380.38]**</td>
<td>[1.983]**</td>
</tr>
<tr>
<td>Durbin-Watson</td>
<td>2.45511</td>
<td>1.796</td>
<td>2.45511</td>
<td>1.796</td>
<td></td>
</tr>
<tr>
<td>Runs Test Obs. #</td>
<td>11.</td>
<td>Not Significant</td>
<td>18.</td>
<td>22.</td>
<td>45.</td>
</tr>
<tr>
<td>Exp. #</td>
<td>18.</td>
<td>18.9</td>
<td>40.</td>
<td>-4.</td>
<td>42.</td>
</tr>
<tr>
<td>Runs Ratio</td>
<td>-3.5*</td>
<td>-3.228</td>
<td>.0972</td>
<td>-.379</td>
<td></td>
</tr>
<tr>
<td>Autocorrelation (lag 1) of the Residuals</td>
<td>.0496</td>
<td>-.228</td>
<td>.0972</td>
<td>-.379</td>
<td></td>
</tr>
<tr>
<td>Autocorrelation of the Residuals in Excess of .3/lag</td>
<td>-3.35/5</td>
<td>-3.22/9</td>
<td>22.</td>
<td>42.</td>
<td></td>
</tr>
<tr>
<td>Kolmogorov-Smirnov Statistic</td>
<td>.1291</td>
<td>.1892</td>
<td>.167</td>
<td>.221</td>
<td></td>
</tr>
<tr>
<td>Residuals: Skewness Kurtosis</td>
<td>-1.597</td>
<td>.773</td>
<td>-4.583</td>
<td>.055</td>
<td></td>
</tr>
<tr>
<td>1st Quarter January</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>February March</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>April May</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>June</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td>None</td>
<td></td>
</tr>
</tbody>
</table>
## TABLE F
Regression Model of Allowance for Borrowed Funds Used During Construction

<table>
<thead>
<tr>
<th>Independent Variables:</th>
<th>MODELS</th>
<th>MODELS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient (Standard Error)</td>
<td>Levels</td>
<td>First Differences</td>
</tr>
<tr>
<td>t-Statistic</td>
<td>36 obs.</td>
<td>35 obs.</td>
</tr>
<tr>
<td>Constant</td>
<td>2217.85 (3617.29)</td>
<td>184.92 (133.32)</td>
</tr>
<tr>
<td></td>
<td>.6131</td>
<td>1.387</td>
</tr>
<tr>
<td>Budgeted Allowance BBFIC</td>
<td>.3309 (.215)</td>
<td>.57992 (.2717)</td>
</tr>
<tr>
<td></td>
<td>1.558</td>
<td>2.135*</td>
</tr>
<tr>
<td>CPI</td>
<td>-9.163 (22.26)</td>
<td>-11.239 (70.133)</td>
</tr>
<tr>
<td></td>
<td>-.406</td>
<td>-.360</td>
</tr>
<tr>
<td>Construction Work-in-Progress</td>
<td>.0041 (.0035)</td>
<td>.0162 (.0062)</td>
</tr>
<tr>
<td></td>
<td>1.23</td>
<td>1.641</td>
</tr>
<tr>
<td>Interest Long-Term Debt</td>
<td>.6705 (.243)</td>
<td>.218 (.302)</td>
</tr>
<tr>
<td></td>
<td>2.765**</td>
<td>.722</td>
</tr>
<tr>
<td>Long-Term Mortgage Bonds</td>
<td>-.0012 (.0018)</td>
<td>-.002 (.002)</td>
</tr>
<tr>
<td></td>
<td>-1.8115</td>
<td>-1.686</td>
</tr>
<tr>
<td>Dividends—Preferred Stock</td>
<td>-.5532 (.3975)</td>
<td>-.764 (.358)</td>
</tr>
<tr>
<td></td>
<td>-.1393</td>
<td>-.172</td>
</tr>
<tr>
<td>Interest Rate. Utility Preferred Stock of Medium Grade</td>
<td>.0928 (1.197)</td>
<td>-.073 (.835)</td>
</tr>
<tr>
<td></td>
<td>.0631</td>
<td>-.0398</td>
</tr>
<tr>
<td>Model Precision at 95% Confidence [i.e., Standard Error x (1.96)]</td>
<td>332.098</td>
<td>447.80</td>
</tr>
<tr>
<td>R²</td>
<td>.969</td>
<td>.299</td>
</tr>
<tr>
<td>(R²)</td>
<td>(.961)</td>
<td>(.148)</td>
</tr>
<tr>
<td>[F-Ratio]</td>
<td>[125.33]**</td>
<td>[1.987]**</td>
</tr>
<tr>
<td>Durbin-Watson</td>
<td>2.2291</td>
<td>2.7681</td>
</tr>
<tr>
<td>Runs Test</td>
<td>Not Significant</td>
<td>Not Significant</td>
</tr>
<tr>
<td>Obs. 1</td>
<td>20.</td>
<td>21.</td>
</tr>
<tr>
<td>Exp. 1</td>
<td>18.9</td>
<td>18.4</td>
</tr>
<tr>
<td>Runs Ratio</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Autocorrelation (lag 1) of the Residuals</td>
<td>-.1414</td>
<td>-.393</td>
</tr>
<tr>
<td>Autocorrelation of the Residuals in Excess of .3/lag</td>
<td>--</td>
<td>-.370/3</td>
</tr>
<tr>
<td>Kolmogorov-Smirnov Statistic</td>
<td>.1609</td>
<td>.1877</td>
</tr>
<tr>
<td>Residuals: Skewness Kurtosis</td>
<td>.5078 (2.947)</td>
<td>.4192 (4.274)</td>
</tr>
<tr>
<td></td>
<td>5.3647</td>
<td>5.826</td>
</tr>
<tr>
<td>Outliers 1st Quarter</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>January</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>February</td>
<td>Below 301.22</td>
<td>None</td>
</tr>
<tr>
<td>March</td>
<td>Below 179.81</td>
<td>None</td>
</tr>
<tr>
<td>2nd Quarter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>April</td>
<td></td>
<td></td>
</tr>
<tr>
<td>May</td>
<td></td>
<td></td>
</tr>
<tr>
<td>June</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Independent Variables:</td>
<td>Coefficient (Standard Error)</td>
<td>Models</td>
</tr>
<tr>
<td>------------------------</td>
<td>-------------------------------</td>
<td>--------</td>
</tr>
<tr>
<td></td>
<td>Levels 36 obs. SELECTED</td>
<td>First Differences 35 obs.</td>
</tr>
<tr>
<td><strong>Constant</strong></td>
<td>517.11 (895.65)</td>
<td>70.92 (59.66)</td>
</tr>
<tr>
<td><strong>CPI</strong></td>
<td>-6.584 (6.457)</td>
<td>-44.55 (36.40)</td>
</tr>
<tr>
<td><strong>Budgeted Interest--LTD</strong></td>
<td>0.7802 (.117)</td>
<td>0.472 (.153)</td>
</tr>
<tr>
<td><strong>Allowance for Borrowed Funds During Construction</strong></td>
<td>0.051 (.107)</td>
<td>0.036 (.092)</td>
</tr>
<tr>
<td><strong>Net Income</strong></td>
<td>-0.006 (.004)</td>
<td>0.000 (.006)</td>
</tr>
<tr>
<td><strong>Dividends--Preferred Stock</strong></td>
<td>0.066 (.205)</td>
<td>0.059 (.300)</td>
</tr>
<tr>
<td><strong>Utility Bond Rates (Baa Rating)</strong></td>
<td>0.450 (.58.25)</td>
<td>278.76 (24.84)</td>
</tr>
<tr>
<td><strong>Model Precision at 95% Confidence [i.e., Standard Error x (1.96)]</strong></td>
<td>212.03</td>
<td>237.07</td>
</tr>
<tr>
<td><strong>Auditor's Desired Precision</strong></td>
<td>450.</td>
<td>450.</td>
</tr>
<tr>
<td><strong>R²</strong></td>
<td>.9662 (.959)</td>
<td>.427 (.304)</td>
</tr>
<tr>
<td><strong>[F-Ratio]</strong></td>
<td>[138.13]**</td>
<td>[5.46]**</td>
</tr>
<tr>
<td><strong>Durbin-Watson</strong></td>
<td>1.481</td>
<td>2.059</td>
</tr>
<tr>
<td><strong>Exp. 19.</strong></td>
<td>Not Significant</td>
<td>42.</td>
</tr>
<tr>
<td><strong>Runs Ratio</strong></td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td><strong>Autocorrelation (lag 1) of the Residuals</strong></td>
<td>.1864</td>
<td>.183</td>
</tr>
<tr>
<td><strong>Autocorrelation of the Residuals in Excess of .3/lag</strong></td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td><strong>Kolmogorov-Smirnov Statistic</strong></td>
<td>.0889</td>
<td>.125</td>
</tr>
<tr>
<td><strong>Residuals: Skewness Kurtosis</strong></td>
<td>.345</td>
<td>.762</td>
</tr>
<tr>
<td><strong>Outliers</strong></td>
<td>1st Quarter Below 57.04</td>
<td>Below 280.31</td>
</tr>
<tr>
<td></td>
<td>January Below 59.38</td>
<td>Below 1015.7</td>
</tr>
<tr>
<td></td>
<td>February Below 104.5</td>
<td>Below 482.89</td>
</tr>
<tr>
<td></td>
<td>March Below 59.38</td>
<td>Below 127.22</td>
</tr>
<tr>
<td></td>
<td>April Below 1001.7</td>
<td>Below 107.71</td>
</tr>
<tr>
<td></td>
<td>May Above 611.17</td>
<td>Above 597.32</td>
</tr>
<tr>
<td></td>
<td>June Above 611.17</td>
<td>Above 597.32</td>
</tr>
</tbody>
</table>
The Application of Regression Analysis for Limited Review and Audit Planning:
A Discussion

by

Andrew D. Bailey, Jr.

University of Minnesota
The Application of Regression Analysis for Limited Review and Audit Planning: A Discussion

Introduction

The authors of this study wish to test the effectiveness of regression as an analytic review tool in an audit setting. The measure of effectiveness employed is a comparison of the regression results to traditional analytic review techniques. As traditional analytic review techniques are highly judgement based, this choice of comparison forces the authors into a field study setting. Field studies of this type provide unique benefits to the researcher, but only at a certain cost. This cost is reflected in the generally ad hoc nature of explanations with respect to observed study results. Unlike a simulation study, the researchers in a field study can seldom be certain that their explanations are based on a relevant and unchanging environmental specification. Of course while a simulation study can establish truth for comparison purposes, its truth is artificial and generalization from it to the "real world" questionable. But, how valid are generalizations from a "real world" case study likely to be when applied to a different setting?

Whatever your answer to the above musings it would seem that field application papers of this type are a necessary part of the often long and difficult process of translating and demonstrating the usefulness of research results for practice. This paper attempts to examine the relevant results in statistics and auditing theory for a significant practical audit problem.

The balance of this discussion is presented in two parts. First a once through overview of the major conceptual issues is presented. This is followed by a more detailed discussion of a number of specific issues presented in the paper. I admit that much of what I have to say was anticipated by the authors in their discussion of the limitations of their own work. In many
cases my comments are for clarification or to provide emphasis rather than to criticize.

Method of Study

It may be worthwhile to point out something the authors do not intend to do and thus cannot be criticized for a failure to accomplish. This paper does not intend to introduce new or novel mathematical or auditing techniques. Thus, any criticism must be limited to the quality of application of these generally well known methods. In this respect it is obvious that the authors are careful practitioners of mathematics and auditing.

The contribution of the study to new knowledge derives from the empirical validity of the application. The researchers employ auditors in an audit setting to test the effectiveness of a linear regression methodology as compared to traditional analytic review techniques.

Regardless of the technology applied, analytic review involves: predicting an outcome such as an account balance; comparing the prediction to observed conditions, e.g. the reported account balance; making a judgement as to the significance of any observed discrepancy; and finally making a decision as to the appropriateness of a search if the discrepancy is judged to be significant. This study of analytic review techniques makes comparisons as to the effectiveness of two competing analytic review tools. Two basic measures of effectiveness were available in comparing the model alternatives. The intramodel predictions, judgements and searches can be studied or a direct measure of the success of the competing techniques in making correct predictions can be assessed. The second noted alternative measure of effectiveness is often employed in simulations where the environment is clearly specified and controlled, but is seldom even seriously pursued in field studies. The problems
of such an attempt are varied, but generally involve the cost of a complete
search of all population observations and the inability of the researcher to
objectively define the truth of their observations and explanations. In this
study every account would have to be analyzed in order to establish the degree
of predictive accuracy present in the two techniques. Without a theoretic
causal model, the search of the accounts in this fashion would be subject
to the criticism that the explanations were ad hoc. The authors, not
unexpectedly, choose not to follow this approach, but rather to make only
intramodel comparisons, investigating only those discrepancies identified by
the alternate methods under study. This choice is obviously limited by the
absence of a base-line comparison. It can, however, potentially rank the
effectiveness of the alternatives. Such a ranking is unambiguous only for
the case wherein one of the methods identifies a subset of the outliers
identified by the alternative method. The authors seem to have happened upon
this gratuitous set of circumstances. They of course remain subject to the
criticism of the ex post and ad hoc nature of their explanations and results.

The Choice of a Regression Application

In choosing regression the authors imply a preference as to its likely
appropriateness as compared to other available approaches. The authors
may have chosen this method only because others had suggested it, a sufficient
reason for a field study. Yet, in their attempt to justify the approach
selected they explicitly commented on an alternative involving simultaneous
equations. They suggested that the simultaneous equations approach cannot
be supported because it is dependent upon a nonexistent "clear theory of the
relationships between multiple equations." Elsewhere they suggest that their
approach provides a "well-specified theory-based relationship." These comments when taken out of context may seem unjustly emphasized. However, a clear drift toward the assertion that the prespecified regression models are more theoretically sound than the simultaneous equation models is obvious in the paper. Both approaches lack a complete and cohesive theory justification. It seems to me that the simultaneous equation models can only be judged in the same manner the authors propose for those models considered in this study.

I concur with the authors, who state that the simultaneous equations approach does "permit the adjustment of the correlation between explanatory variables and the error term." A similar benefit might have been obtained by the authors in this study were they to apply the technique of seemingly unrelated regressions.

The authors also reject a stepwise regression approach to the selection of variables. Their concerns in this case were mathematical and legal. I accept that the stepwise method does introduce certain mathematical difficulties which include overspecification of the model and problems of statistical interpretation problems. I am somewhat confused however by their apparent emphasis of the legal issues. These issues are as yet untested and it seems to me might cut both ways. For example, the authors are concerned with being able to justify the particular variables selected. If the set of variables originally identified for stepwise consideration is supportable as potentially explanatory, why couldn't the auditor explain the selection of any subset as well as any other subset? If multicolinearity is the problem leading to exclusion of some variables that might be of legal interest, inclusion of the correlated variable will not invalidate the prediction. Limited degrees of freedom may of course be a problem. As suggested by the authors if "a well specified theory-based relationship" exists or even a partial theory, certain variables might be excluded on the basis of the signs of the coefficients.
The authors have failed to identify these relationships in their paper.

My second concern with relying on an untested legal issue is in the risk that the jury may not understand the sophisticated, although valid, statistical arguments that distinguish the analytic selection process of stepwise regression from the auditor heuristic processes of model specification, search for outlier explanation and model augmentation. As noted by the authors the models may be augmented based on the ad hoc explanations obtained from an incompletely prespecified initial model. How will one respond to the sharp prosecutor who states that: "We note that you used a prespecified regression model in your analytic review procedures. That model includes only a part of the variables set available to you. My office in applying stepwise regression notes that a 'better' predictive model can be developed from the data available to you at the time-of-audit. Had our model been employed the account now in question would have been investigated and this fraud uncovered on a timely basis." I expect that the authors will provide solid technical and conceptual argument, but as they observe the issue is now legal, not technical. It is not clear to me at this time that their case will prevail. Collins has some interesting observations on the choice of methods and decision rules that relate to this very matter.

Model Limitations

The choice of regression as a technique leads to the authors' expenditure of significant amounts of effort in model specification and testing. The authors are very thorough in establishing the correspondence of the model assumptions and data input. I would, however, like to take a moment to reemphasize several issues. First, to obtain a data set of 84 observations, the authors have relied upon unaudited data. In doing so the authors rightly
point out that they do not have the same level of assurance that the data are free of measurement error as would exist when using audited data. In assuming "excellent client controls [based on a]...recent study and test of the client's entire information system..." the authors accept that such a study is sufficient to provide reasonable assurance that the monthly data is accurate. Without special attention given to such matters as monthly cut-offs and allocations, this is a risky assumption.

A second issue considered by the authors and of significant concern to me is model stability. The fact that 36 data point models tend to dominate the 84 point data models suggests a problem of underlying instability in the environment. Beyond the question of the basic stability within the period over which data was collected, is the question of utilizing new data observations for prediction purposes when the new data observations are outside the range of all prior observational data. I expect this to be the case for a number of the independent variable used in this study, e.g. CPI indices, budget data, MDF and KWH data.

Comment

The authors have provided a rather complete list of concerns regarding model limitations. The above comments, those preceding them, and those of the authors are not fatal. They are of concern to me in part because of the authors' somewhat offhand manner of setting aside techniques other than regression.

Analytic Investigation of Residuals

As noted above, this study is limited in its ability to assess the effectiveness of any analytic review technique since it only investigates
outliers identified by the models under study. The study can provide no
direct evidence as to whether other accounts should have been of concern as
well as those identified. Further, the explanations obtained in a real audit
setting are themselves only potentially acceptable in that they may or may not
be the complete story.

Using only the identified outliers, the authors compare four approaches
to the selection of accounts for investigation. Implicit in the comparison
is a cost/benefit analysis of the specific account selection as well as the
general methodology. It is disappointing, but understandable, that the
authors have not explicitly introduced cost/benefit and auditor utility issues.
Without the introduction of these factors it is not possible to conclude that
one search technique is preferred over another. It is the case that some
strategies lead to more investigation than others and it is unlikely that
investigating every outlier in depth is worth the cost. A more complete
statement of the auditor utilities, cost and benefits is necessary to resolve
this issue.

Miscellaneous Questions

1) Desired precision has been set at 1/2 the monthly materiality cut-off
per account. Is the materiality cut-off per month per account equal to 1/12
the yearly materiality measure? Is this appropriate?

2) Materiality has been associated with the auditors' degree of responsibility.
It would seem that the degree of confidence is more appropriately associated
with the degree of responsibility. Is materiality a function of the degree
of responsibility as suggested by the authors?

3) Can data within the firm be neatly categorized as internal to the
accounting function and external to the accounting function given the common
data base involved?
4) Can the concept of risk exposure provide a basis for formulating regression models in any formalized way or is it purely by example as employed by the authors?

5) Does the authors' suggestion that "a careful definition of endogeneous... permits the auditor to formulate a reduced-form equation with only exogeneous variables..." presume more knowledge with respect to the accounting control system than is generally available?

6) Is it possible for the authors to create a more readable table?

7) Could the authors create a more controlled test with respect to managers model selection as a substitute for the traditional analytic review methods?

Conclusion

This paper presents an interesting, but incomplete analysis of an application of regression in an audit setting. It is not definitive in its measures of effectiveness, but weakly suggests that linear models may be a substitute for man in identifying accounts for potential further audit study or in-depth examination.
The Application of Regression Analysis
for Limited Review and Audit Planning:
A Discussion

by
William C. Mair
Touche Ross & Co.
GENERAL SUMMARY:

The case presented here represents a single instance demonstrating that regression analysis based on judgmental identification of determinant factors can be reasonably effective and efficient. It demonstrates that it can be done by auditors, but does not provide much guidance as to when else these techniques will be appropriate.

My comments will consist first of some detailed remarks on specific statements in the paper, and then general remarks regarding the methods and findings described in the paper. Some points in the paper have been clarified by the authors' verbal presentation.

DETAILED COMMENTS:

The example industry may be atypically ideal for the field experiment. This is not objectionable per se, as it would obviously not make sense to attempt a clearly impossible situation. It does, however, bring highly limited transferability, because the country contains relatively few gas and electric utilities.

The example industry may be atypical both because of its size and in the predictability of its revenues. For example, as a consumer, I am aware that my monthly utility bill is frequently annotated "estimated." To the extent that this occurs frequently, and further to the extent that the same estimators are used for the billing estimate as by the auditor, the audit forecasts might not be as independent as they appear.
The comment on Page 71 that the training for both statistics and regression analysis took only 1-1/2 days is quite impressive. However, this seems to be contradicted on page 110 which states that attribute and variables estimation training required 3 days and regression analysis required an additional 2 - 3 days (total 5 - 6 days).

I think I rather resent the gratuitous comment on Page 104 regarding the resistance of auditors to change. I find that successful auditors are frequently quite different from this stereotype. I know personally that I am prone to "change for changes sake." I have learned to listen to those "conservative auditors." If my proposals meet such resistance, I have learned to reexamine my proposals, because those "conservative auditors" might have a better understanding of the "real world" than I do. This case admits to substantial limitations on the applicability of quantitative forecasting methods. The appropriate reaction to rejection by learned and experience persons is to question why, not dogmatism of another type.

The example on Page 76 regarding segregation of duties seem to have little relevance. We can rarely obtain positive reliance on segregation of duties, and "manipulation" is seldom a material exposure. The greatest thief can rarely surpass a moderately accomplished "incompetent."

I'm somewhat curious regarding which timesharing facilities on Page 80 were considered and rejected for the reasons given.

Even after several readings of the paper, certain of the findings still remain unclear. The paper states that "no significant events took place which were not identified by the regression formulas." A more true summary of the facts that are represented might have been, "the regression formulas identified all significant events also identified by conventional review."

However, were any significant events found by the regression formulas that were not found in formal review? There is an indication on page 96 that
the regression model found a policy shift in long-term financing that might not have been previously known, but the implication of this is never discussed elsewhere.

Some of the cost estimates also lead to additional questions. A statement is made that overall audit hours might decline by 5%. However, this generalization doesn't disclose the basis for the estimate, or whether the proportion is applicable to clients of different sizes or industries.

The consequences of replacing the auditor model with a stepwise regression also needs some clarification. Page 108 indicates that the stepwise model found all of the outliers found by the auditor model, plus explained four more. Were these four not found by the auditor model, or where they found by both judgmental review and the auditor model, but resolved, and not considered to be outliers by the stepwise model?

Finally, the readability of the entire paper could be improved by consistently defining all abbreviations, symbols, and obscure (at least to me) mathematical terms.

GENERAL REMARKS:

The paper addresses the application of regression techniques to four potential audit uses:

- a limited review, as in the case presented
- identify "problem areas"
- help define audit scope
- perform a final review of explanations

Perhaps other uses are not germane to the issues, but additional uses of this or similar methods would include:

- Estimates of future obligations, such as warranty or casualty insurance payments. Logarithmic, pure time series, or difference models
would probably work best.

-Cash-flow forecasts, particularly as they might influence a "going concern" opinion.

-Reasonableness tests of monthly operating items after one month was tested in detail, so as to avoid testing of "transactions throughout the period under examination." For example, we sometimes parallel simulate a single month's mortgage revenue and then compare the yield in the tested month to the revenue for the remaining eleven months.

-Reports on forecasts, as described in a very recent Audit Guide.

-Strategic consulting, as "what if?"

I am pleased, an auditor, with the rationale for using the methods described, and for the concern exercised over the potential pitfalls.

The auditor-specified models imply a reliance on known economic factors, which should prevent serious omission of structural variables, spurious correlations, or some causes of excessive standard deviation. The statistical tests should detect non-linear relationships, changes or shifts in relationships, and non-normal differences - all of which I suspect are relatively common. Finally, the evaluation of internal control should minimize risks from erroneous input data.

My interpretation from reading the new Forecasting Audit Guide and SAS #23 on analytical reviews is that forecasting models must include all key determinants that may materially affect the estimate. This seems to support the professional standards applied in this case, and probably precludes the use of univariate time series regressions.

I remain a bit skeptical, however, regarding the general incorporation of budget, internal forecast, or billing data determined by forecasts in auditor forecasts. The use may have been suitable in this specific case, but generally seems to be prone to self-fulfilling prophecies.
In this regard, and in light of the discussion on pages 76 and 77 regarding the relative quality of evidence, I wonder whether an explicit weighing factor should be assigned to such situations.

If I may summarize the specific costs discussed on page 110 I estimate that first year application costs totalled approximately $5,000, and costs for training and client understanding are additional. Savings of this specific investment, plus 5% of total time charges, implies fairly sizable time charges before savings of the magnitude noted should be expected.

If one of the underlying objectives of this paper was to what my interest in regression analysis techniques, it succeeded. While the case includes relatively few circumstances that would cause one to expect widespread utility, it does demonstrate that practical use is feasible, economical, and reliable, in at least some circumstances. These circumstances might be assumed to include:

-A good understanding of the client environment (which might limit its application on new clients)
- Stable relationships between determinants
- Good internal control over determinant data
- Sizable level of audit time charges

This finding comes as no special surprise, since several people had predicted this and one major accounting firm has been a strong field proponent.

I'm also reassured that "reactionary auditors" can sometimes do (nearly) as well as mathematical methods, without losing much sleep over "heteroscedasticity.

Given that regression analysis can work, but its use is optional and influenced heavily by effects on audit costs, the most important further research would relate to the elements of costs and methods to impact them.
My hope would be that, even if the total number of hours required increased, we could shift those hours toward less experienced, lower cost personnel. However this important precondition carries significant cost, which has not been mentioned.

CONCLUSION:

The primary benefit of regression analysis appears to be that it forces discipline in the review and documents the rationale.

The next priority should be to search for method for:

- Screening for appropriate environments,
- efficiently coping with non-ideal environments, and
- limiting requirements for prior knowledge, such as industry parameters, that could reduce the demands for internal control evaluation, prior client experience, and prior paractice experience.

I encourage the use of the "scientific method" of observing the environment and attempting to fit mathematical models. Continued experiments should be pursued. I agree that the experiment described in this paper is a step in the right direction.
Addendum to: "The Application of Regression Analysis for Limited Review and Audit Planning"

by
Abraham D. Akresh
and
Wanda A. Wallace
ADDENDUM TO: "The Application of Regression Analysis for Limited Review and Audit Planning"

Participants at the Symposium requested that the authors' verbal comments and the text of the significant questions and answers from the session be included in the published proceedings; in response to that request, this addendum was prepared.

I. SYMPOSIUM COMMENTS BY ABRAHAM D. AKRESH

Professor Wanda Wallace and I wrote this paper to describe an application of regression analysis for limited review and audit planning. In our presentation, I will explain the history of PW's involvement with regression analysis and what we were attempting to accomplish. Then, I will discuss what I consider the advantages of regression analysis for the auditor. After, that Professor Wallace will discuss some of the technical issues and point out some solutions we reached. We are extremely interested in the reaction of the discussants and also intend to leave plenty of time for the audience to ask questions.

History of This Project

PW has been interested in regression for at least the past three years. We were one of the firms in the regression analysis joint venture which led to the two case studies on regression analysis for analytical review that John Neter will be discussing this afternoon. However, we felt the need to perform some study on our own.

Our interest in regression increased at the 1978 Illinois Conference when I read some of the papers implying that the method might work. In 1979, we
decided to try a project in the area of regression analysis. PW hired Professor Wallace, first as a Faculty Fellow and then as a consultant, during the summer of 1979. We applied regression analysis to the audit of a public utility. This 1979 project was a predecessor to the project we described in our paper.

Our goal in the 1979 project was to determine whether regression analysis would work; that is, would it yield the same answers as detailed audit testing. Our goal was not to determine whether a staff auditor could use the technique but whether we could actually prove anything with the tool.

Our approach to the 1979 project was to take a relatively low risk client and re-perform the 1978 audit using regression analysis. Professor Wallace reviewed the prior year's working papers, set up the models, and determined what the outliers were. We then compared the results with what the auditors had previously determined. In all cases, regression found what the auditors had found and, more importantly, what the auditors found in the audit would have been detected by regression.

This experiment was designed to strengthen our belief in regression and teach us some of its pitfalls. We learned that, indeed, a consultant could set up a useful regression model which met the auditor's requirements and which could be accomplished within the time constraints of an audit. We also learned, however, that the available time-sharing programs were inadequate for auditor use.

We first tried to use the STAT II system on General Electric. That appeared to have certain advantages because our staff was used to G.E. However, G.E. did not have all of the necessary computations on its system. Thus, we performed certain computations using the IDA package. Still other computations
required additional software from the University of Rochester. In short, performing the analysis was a time consuming procedure, requiring the use of several software packages and subroutines.

With these results from the 1979 project, we concluded that if we were to move forward in regression, we would need two things:

(1) An auditor oriented time-sharing system; and

(2) An appropriate continuing education program.

We first set out to design the time-sharing program in early 1980. By late April, we had the rough workings of a good time-sharing program. We have continued to work with our time-sharing program to put in additional edits and other checks. Like all of our time-sharing programs, our regression program is built on the assumption that the auditor is not a mathematician, and that he need not understand the details of the formulas. Rather, the program checks everything that can be checked, including the necessary statistical tests.

In addition to the time-sharing program, we developed educational material. Our material consisted of a 3-hour overview and a 12-hour detailed course. This course is designed to permit the auditor to understand regression analysis and apply it by using the PW time-sharing program.

The course was initially taught by Professor Wallace to 20 auditors who had previously completed the firm's statistical sampling training programs. In the future, it is expected that the scope of the regression analysis course will be expanded. After the course, several auditors have attempted to use regression analysis. These applications are still in progress.

The 1980 study, which is described in the paper presented, involved a limited review. It differs from the 1979 study because the audit manager had
attended our regression analysis training program and had learned how to
use the PW time-sharing system. In the 1980 study, all of the relationships
were specified by the audit manager. Most of the running of the time-sharing
program was done by the audit team. The audit manager also specified the
desired materiality and evaluated the results. Finally, the audit manager
determined the reasonableness of any outliers.

The 1979 study was designed to determine whether regression would work
when set up by a consultant. The 1980 study was designed for a different
purpose—to determine whether an auditor with much less interaction with a
consultant could perform regression analysis. I am firmly convinced that
auditors can be taught to effectively perform regression analysis in an
auditing context.

We were also looking to see if we could put some meaning into the limited
review process. The limited review relies heavily on analytical review, and
we believed that regression would be well suited to limited reviews.

In this study, we took the same conservative approach we took in 1979;
that is, we performed the review both ways—first in the traditional manner
and then using regression analysis. In the limited reviews for the first
two quarters, we performed the traditional limited review before using re-
gression. Then the audit manager, prior to reviewing the working papers and
after completing the 1 1/2 day training session, specified the accounts to be
analyzed and the expected regression relationships. We compared results; in
all cases regression analysis found whatever traditional procedures found.
In several cases, regression raised additional considerations which had been
missed by traditional procedures. For the third quarter limited review, we
will be using regression on a "live" basis.
Advantage of Regression Analysis

I would like to tell you what I think are some of the practical benefits of regression analysis, either in a limited review or in an audit.

• First, in the limited review and in the audit, we obtain a much more disciplined approach to analytical review. Analytical review becomes a procedure with an objective and a rationale. It is no longer a once over lightly procedure. Regression, or any disciplined mathematical procedure, forces the auditor to think, and when the auditor thinks, he audits smarter.

• Second, regression forces the auditor to document. We have developed forms to help the auditor through the regression logic. By using these forms, the auditor is forced to specify his objectives, his variables, how he expects them to change, what materiality is, and what an outlier is. In short, regression is bringing to analytical review what statistical sampling brought to audit testing.

• Third, regression forces the auditor to really understand the business, not just check the numbers. To effectively perform regression, the auditor needs to understand business variables. For example, to predict utility revenue, he has to understand the production cycle, the relationship between revenue and temperature, the relationship between revenue and prices, and various other factors. He also has to find information outside of the accounting records; for example, he must go into the production department, perhaps into the rate-making department, and so forth. He also must go outside of the company to obtain industry data. This gives him a better understanding of the financial statements and makes him more useful to client management. Thus, the management letter often becomes more meaningful.

• Fourth, in the long run, time will be saved. The auditor does not have to spend time tracking down and explaining variations which are not variations but amounts that are actually in line when all the variables are considered. In the traditional approach, if sales revenue goes up 15%, the auditor needs to find an explanation. In regression analysis, the auditor's model may tell him that the increase in sales of 15% is normal when the auditor considers the independent variables of production, the consumer price index, and degree days. Thus, after the model is built, it is possible that the auditor will not have any explanation in the working papers for changes because all of the variables will have been incorporated into the model and thus, not be unusual.
Obviously, we have a great deal to do before regression is an important procedure in most audits and limited reviews. But, we do believe that this paper represents a step in the right direction. Professor Wallace next comments on some of the technical issues which we addressed, and how we solved some of them.

II. SYMPOSIUM COMMENTS BY WANDA A. WALLACE

There were several critical decisions made in the planning stage of this field experiment. One key question concerned what type of data to utilize in regression model building

(a) external data
(b) budgeted data
(c) internal data
(d) or some combination of internal and external data.

Another issue concerned whether to utilize a general accounting model or a client-tailored regression model. The factors analyzed in addressing these questions included

• endogeneity issues or, in other words, the results of the internal control evaluation,
• the extent to which simultaneous equations were feasible, and
• the consequent trade-offs in precision (as exemplified in Table 1 of our paper).

The decision was to utilize client specific reduced-form regression models which integrated internal and external data.

A third question was whether to utilize existing time-sharing regression packages or to develop a package tailored to the audit setting. Due to the numerous problems cited in our manuscript, PW & Co. opted to develop its own software package. At this point, I would like to highlight the novel aspects of the software design specifications:
(1) Automatic statistical checks, directed at known problems with accounting time series data are provided.

(2) The auditor is directed in selecting appropriate data transformations to address any deficiencies identified by the program, based on the auditors' reply to questions concerning the viability of a multiplicative or reciprocal relationship.

(3) The auditor is warned of model limitations.

(4) Most of the terminology utilized in the output is already understood by auditors familiar with statistical sampling.

(5) The usefulness of first-difference models in checking model specification is explicitly incorporated in the package.

(6) The use of larger sample sizes is assumed, with Chow model shift tests and standard error comparisons leading to reduction of time-series samples when preferred.

(7) A stepwise selection procedure utilizing an all possible regressions (or optimum regression) test on the selected number of regressors is available, although its use as the primary means of model specification is discouraged.

(8) Simultaneous prediction intervals are utilized.

(9) A comparison of the model's achieved precision (defined as $1/2$ the $95\%$ prediction interval) and desired precision is highlighted in the output.

(10) The amount by which specific monthly book values lie outside the lower or upper bounds on predictions is reported.

The time-sharing package is conservative in the sense that numerous statistical checks are performed, some of which are redundant and some of which will probably appear significant simply due to chance, since each data set is subjected to numerous tests. To provide some preliminary indication of how likely audit models are to fail the statistical checks and therefore to require transformation, Table 2 was formulated. Despite some problems in the various models tested, it was possible via changes in the
number of observations utilized for the base model or the form of the model utilized (first-differences versus levels), to derive a set of technically valid base models with relative ease.

Having assessed the statistical propriety of the regression models, the audit usefulness of these models was considered. Not only were similar fluctuations noted in the regression application and traditional analytical review procedures, as Abe has discussed, but regression added a historical perspective of client operations to the traditional analysis. Specifically, insight was gained regarding a shift in the management's financing policy. When the regression models were expanded to check both the reasonableness and sufficiency of explanations for fluctuations which had been noted in the traditional variation analysis, either the outliers were explained, or additional investigation pinpointed problem areas not detected in the traditional review process.

The use of the limited review results in the course of audit planning led to the auditors' reduction of detailed tests in 6 of the 7 accounts reviewed and an estimate of a 5% decline in overall audit hours.

The paper demonstrates two key points:

(1) It supports the ability of auditors to specify regression models with audit resolving power when compared with both
   a. auditors' desired precision and
   b. the benchmark of stepwise-selected models' standard error

   and (2) It further demonstrates how four available methods of computing residual quantities differ in their audit investigation implications.

The experience of PW & Co. supports the unsurprising fact that the training requirements of regression are the most substantial cost factor, while computer costs are nominal—about $3 per run and $300 for initial data file construction (assuming 20 variables).
confronted concerns determining where and when to apply regression; the
pointers provided in the manuscript can be expected to be augmented based
on future field studies.

Overall, much research remains to be done in the area of regression
applications to the audit, as suggested by the rather long list provided in
our manuscript. However, this study suggests substantial savings in audit
time and efficiency in sample selection as well as a more objective analytical
review can result from applying the regression procedure.

III. SIGNIFICANT QUESTIONS AND ANSWERS

**QUESTION:** There seems to have been some inconsistency in your remarks
concerning the training time you think is needed for regression. Please
clarify.

**RESPONSE:** The participants in the training course we presented in April were
seniors and managers who teach statistical sampling at PW. Over the years,
they had attended an introductory seminar of three days, an advanced seminar
of two days, and a two-day instructors' seminar. The regression course for
them was 12 hours and this was not enough.

On an ongoing basis, before taking the regression course, the auditor
needs the three-day introductory seminar, where the basic statistical con-
cepts and how to use timesharing are taught. The regression course itself
should be 2-3 days and should include hands on experience.

**QUESTION:** This seems to be a report of a single example in a highly regulated
industry. I do not believe the results should be generalized to other situations.
Please comment.
RESPONSE: We admit that this is but a sample of one in a field study environment and thus it suffers from the disadvantages of samples of one. We are not recommending that regression be used on every audit and for every account. We agree that regression will be difficult on initial audits and on companies which do not have stable operations. However, we believe regression will be usable in industries such as utilities, transportation, banking, insurance, retailing, etc. It is up to the auditor to determine in the planning stage whether regression is practical.

It is also important to note that even when the auditor decides to use regression, he will not use it for every account in the financial statements. Rather, he will determine those accounts where it will provide useful evidence. In most cases, these will be income statement rather than balance sheet accounts.

QUESTION: Given utilities use estimated billings for record-keeping, the auditors' use of such numbers in model-building raises a question of independence, does it not?

RESPONSE: The auditor must assess the quality and independent status of data included in regression model-building; this judgment should consider both the type of data involved--internal, external or some combination thereof--and endogeneity or internal control concerns, as discussed in the paper. Rather than using estimated billings, actual production statistics or some external variable like the weather could be used; however, if tests of the estimation procedure were performed, and such procedures were deemed reasonable, even estimated billings could be included in an auditor's regression model.
QUESTION: The field study by nature loses control over the environment; however, there is also no control over the experiment reported in this paper, as the auditors performing the regression analysis knew the client from prior years' audits. Please comment.

RESPONSE: The control we speak of in the paper is directed toward the audit manager's lack of knowledge of where "outliers" existed for the quarter under review. However, we wanted the audit manager to understand client operations based on past audits in order to facilitate model specification and to parallel the typical audit application--since for any given year, only a minor percentage of clients are new.

QUESTION: It appears that the regression models were augmented rather arbitrarily, based on client explanations, suggesting that it is difficult to assess how well the auditor's specified model performed or how independent and objective the models are. Please clarify.

RESPONSE: The base models presented were entirely specified by the auditor and are the focus of the analysis of outliers. The base models were expanded to reflect the clients' explanations of observed outliers as a method of checking the reasonableness of such explanations and of checking the extent to which such explanations accounted for the outliers.

QUESTION: Your estimate of the decline in audit hours of 5% is unlikely to be generalizable. Details as to how this 5% estimate was determined are not provided in the manuscript. Please comment.

RESPONSE: The decline in audit hours that can be expected from the use of regression depends on existing audit procedures prior to the regression application, the achieved precision of the auditor-specified regression
models, and the generalizability of Kinney's (1979) results as to the
decrease in sample size which can result from partitioning based on re-
gression analysis. Also, the field experiment involves a fairly large
audit client for which regression savings are likely to be greater than
those observed for smaller audit engagements.

**QUESTION:** It would appear that a 1/2 of yearly materiality emphasis would
be appropriate as the focus in regression applications, rather than an
emphasis on a 95% confidence level. Please comment.

**RESPONSE:** Risk and materiality are reported in the regression output in the
form of achieved precision; however, a priori the materiality reached in re-
gression is uncontrollable. For initial testing of regression applications,
a 95% confidence level has been set as a benchmark to determine the resolving
power of auditor-specified models relative to the typical reliability levels
desired by auditors. However, as noted in other literature, regression which
offers 50% assurance can contribute toward the overall audit reliability de-
spite the need to augment analytical review procedures to reach the desired
95% level.

**QUESTION:** How is the concept of annual materiality translated to monthly
materiality, and what investigation procedure for outliers is integrated in
the time-sharing package?

**RESPONSE:** Essentially annual materiality divided by twelve, but adjusted for
seasonality, is close to the monthly materiality measures utilized in this
application. Currently all outliers that are flagged would be investigated,
although this is an extremely conservative approach since the likelihood of
an error or misstatement which would be material on an annual basis being
divided equally across all twelve months is low and as a fewer number of months is involved in an error, the probability of its detection increases. The testing of alternative investigation approaches, as discussed in the paper, is planned in future field applications.

**QUESTION:** Since a data base management system typically includes all internal data including accounting statistics, the distinction between data generated by the firm external to the accounting function and accounting data is bogus as an indicator of any level of security. Please comment.

**RESPONSE:** This is similar to the question on estimated billings asked earlier, and the answer is the same. A data base management system does not dictate the independence of the data; the critical issue is how the data are created, not how records are kept. The auditor's judgment is required to assess the independence of the data, which data to include in model-building, and the extent of reliance on regression evidence.
HEURISTICS AND BIASES: SOME IMPLICATIONS FOR PROBABILISTIC
INFERENCE IN AUDITING*

by

Gary C. Biddle

and

Edward J. Joyce

Graduate School of Business
University of Chicago

*The experiments reported in this paper were among those funded by a grant from the Peat, Marwick, Mitchell Foundation through its Research Opportunities in Auditing Program. The views expressed herein are those of the authors and do not necessarily reflect the views of the Peat, Marwick, Mitchell Foundation.
I. INTRODUCTION

In a typical engagement, the auditor must assess the reasonableness of a client's financial statements given imperfect and incomplete information. During the course of the engagement, the auditor might ask himself such questions as: What are the chances that this account is materially in error? What are the odds that this audit test will detect a material error? How should these test results be combined into an overall assessment of audit risk?

Clearly, many audit decisions are based on beliefs concerning the likelihood of uncertain events. As audit technology evolves, auditors are increasingly being asked to formally express these beliefs in planning and conducting audits. (SAS 1 Section 320B is one example.)

How well do auditors cope with uncertainty? If auditors are like many other people, they rely to a great extent on rule-of-thumb procedures to reduce cognitive strain in probabilistic inference tasks. These procedures are called heuristics and serve to make complex tasks manageable. To the extent that they enable auditors to make decisions efficiently and in conformance with normative principles, heuristics are adaptive and valuable cognitive tools. To the extent that they lead to decisions which depart from normative principles, however, the cost of the decision errors must be compared to the cost of making decisions via procedures in conformance with normative principles.
The purpose of this paper is to review the results of three experiments designed to test for behavior consistent with the use of heuristics by auditors. The three experiments reviewed here are a fairly representative sample from a population of 23 experiments which we conducted on heuristic use by auditors. The results of most of these experiments are presented in detail in Joyce and Biddle [forthcoming (a), forthcoming (b)] and Biddle and Joyce [1980]. The next section of this paper provides a brief background on judgmental heuristics.

II. HEURISTICS IN PROBABILISTIC JUDGMENT

The empirical research on human judgment has led psychologists to the following general conclusions [Hogarth, 1975, p. 272]: (1) humans have limited information processing capacity [Miller, 1956, Newell and Simon, 1972; Slovic and Lichtenstein 1971], and (2) the nature of the judgmental task determines to a great extent the decision strategies which they employ [Edwards, 1971; Einhorn and Hogarth, 1980; Payne, 1976; Simon and Newell, 1971; Slovic, Fishhoff, and Lichtenstein, 1977]. How do humans cope with complex situations given their limited cognitive capabilities? The literature cited above suggests that humans resort to cognitively tractable decision strategies known as heuristics. These heuristics are cognitively simpler mechanisms than normative models and often result in decisions which are compatible with the latter [Tversky and Kahneman, 1974]. However, because they imply fundamentally different cognitive activity than normative models,
and because they rely on variables not relevant in normative models (or ignore variables which are relevant in normative models), they sometimes lead to systematic decision errors.

Bayes' Theorem is the normative model where the task facing the decision maker is to revise his beliefs upon receipt of new information. Much of the empirical research on human probabilistic judgment has consisted of comparing human judgments in probability revision tasks to the Bayesian optimal response. The evidence indicates that humans process information in a fashion fundamentally different from Bayes' Theorem. "In his evaluation of evidence man...is not Bayesian at all" [Kahneman and Tversky, 1972].

More recent work in the probabilistic judgment area has focused on developing representations of human judgment that are descriptively more powerful than Bayes' Theorem. In a series of papers, Tversky and Kahneman [1971, 1973, 1974, 1977; Kahneman and Tversky, 1972, 1972b, 1973] have formulated and empirically tested three heuristics which they feel provide more accurate descriptions of human judgment in many situations than does Bayes' Theorem. These heuristics are (1) representativeness, (2) anchoring and adjustment, and (3) availability.

Each of the next three sections discusses one of these heuristics and presents the results of an experiment designed to test for its use in audit judgments. The subjects were practicing auditors from Big Eight public accounting firms. All experiments were pilot-tested on experienced auditors and revised, where necessary, before administration.
III. REPRESENTATIVENESS

Representativeness may be characterized as the tendency to judge the probability that item A comes from population B on the basis of the degree to which A is perceived to be similar in essential characteristics to B. The greater the perceived similarity, the greater the adjudged probability that A is a member of B. Suppose, for example, that you were taking a walk with a friend and noticed a very large, very muscular man about 35 years old with a gravelly voice, cauliflower ear, and lousy disposition. Your friend asks "Do you think he's a professional wrestler?" If you used the representativeness heuristic you'd be inclined to answer in the affirmative. Why? Because his appearance is similar to the stereotypical professional wrestler.

While this procedure possesses a certain intuitive appeal, its use can lend to judgments of probability which are inappropriate. The reason for this is, of course, that the sole criterion is perceived similarity. Other normatively important considerations are ignored. One such consideration is the base rate of professional wrestlers in the U.S. population. There are simply very few professional wrestlers. Even if every professional wrestler fit the above description perfectly (which, of course, is untrue) there would still be many times that number of men who fit the description perfectly, but who are not professional wrestlers.

Another normatively important factor ignored by representativeness is the reliability of data. It is well-documented in the auditing literature (e.g., Arens and Loebbecke, 1980, p. 116) that
the informativeness of data is a function of their reliability. Consider the auditor's review of the adequacy of the allowance for uncollectible receivables. If a very large past-due account exists, a report on the credit-worthiness of the customer from an independent third party such as a reputable credit agency should be considered more reliable -- and thus more informative -- than the same report from the client's credit manager. If the auditor employs the representativeness heuristic, however, source reliability would be ignored and the collectibility of the account would be assessed on the basis of whether the report is representative of a customer who is tardy but pays his bills, or representative of a customer who is a deadbeat. The following experiment was designed to test auditors' sensitivity to the reliability of evidence.

Experiment 1

The subjects in Condition A of Experiment were asked to respond to the following problem:

As part of the regular year-end audit of a client—a consumer-electronics wholesaler—you are reviewing the adequacy of the allowance for uncollectible receivables. You prepare an aging schedule of accounts receivable and note a very large account is six months past due. The customer has returned your positive confirmation verifying the client's balance as correct. You know from your experience with this client that approxi-
mately 50% of account balances six months past due are unrecoverable. Assume this single account balance is a material item. It is the controller's opinion that the entire amount will be recovered and there is no need to provide for the loss. You investigate the customer further and get the following description from the client's credit manager:

The customer is a rapidly expanding merchandiser of television, radio, stereo and other consumer-electronics equipment. It began as a single-store operation in 1974 and now operates a total of 12 stores in three states. Further expansion is planned in the near future. Earnings growth has been strong since 1974. As the firm expanded, its average payment time on accounts receivable has steadily increased. This is due to an inadequate accounting system rather than to cash difficulties. A new computerized accounting system is presently being installed and is expected to remedy the firm's payments problems.

1. Based on the above information, what is your estimate of the probability that the receivable will be collected in full next year? (Circle one number.)

<table>
<thead>
<tr>
<th></th>
<th>.1</th>
<th>.2</th>
<th>.3</th>
<th>.4</th>
<th>.5</th>
<th>.6</th>
<th>.7</th>
<th>.8</th>
<th>.9</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2. Based on the above information, what percentage (if any) of this receivable should be included in the account for uncollectible receivables for this specific customer? ___% 

Condition B was identical except that the subjects were told that the customer description was provided by "an independent credit agency which has provided you with reliable and up-to-date information in the past." Thus the difference between the experimental conditions was the source of the customer description: client's credit manager versus an independent credit agency. In this case, the report of an informed, independent third party should be considered more reliable than the same report made by the client's credit manager. The presumption is that the client's credit manager has more incentive to provide a distorted report than does the outsider.
This problem provides the subjects with two primary types of data: (1) the base rate of uncollectibles — approximately 50 percent for account balances six months past due; and (2) the customer description which does not vary between sources. The customer's description was designed to be representative of a customer who, although tardy, would pay off the entire account balance. Regardless of the source, however, the customer description was not conclusive in this regard (i.e., the customer description was not perfectly diagnostic of collectibility). Thus the base rate data were normatively relevant, and the subjects' predictions of probability of collection in full (part 1) should be regressed toward the base rate of 50 percent. The magnitude of the regression should be contingent, though, on the reliability of the source of the customer description. If the credit manager is a less reliable source than the credit agency for data of the nature contained in the customer's description, the subjects in the credit manager condition should regress their predictions more toward the base rate than subjects in the credit agency condition. Because the greater reliability of the credit agency makes the individuating data more diagnostic, subjects in this condition should make more extreme predictions (i.e., closer to 1.0) than subjects in the credit manager condition.

In part 2 of the experiment, the subjects were asked to estimate the magnitude of the allowance for uncollectibles that should be established for the specific account in question. A subject's response to part 2 should be related to his (her)
response to part 1. That is, the higher the probability of collection in full assessed in part 1, the lower, on average, should be the provision for uncollectibles in part 2.

Subjects. The subjects were 50 practicing auditors from the Chicago offices of two Big Eight public accounting firms. Twenty-seven auditors from Firm 1 were tested first in a single group at the firm's Chicago office. Shortly thereafter 23 auditors from Firm 2 were tested at that firm's Chicago office. These subjects varied widely in their experience. The overall mean experience was 4.00 years with a range from 0 to 27 years. The mean experience level varied considerably between the firms: 2.44 years for Firm 1 and 5.33 years for Firm 2. The present problem was one of 11 problems on probabilistic judgment the subjects were asked to complete. The problems were self-paced and completed in 35-65 minutes. All subjects were randomly assigned to the experimental conditions.

Results. The cell means by experimental condition for both parts of the experiment are reported in Figure 1. The cell means for Part 1 (probability of collection) were virtually identical. They were not significantly different (t = .03, two-tailed). Similarly, the cell means for Part 2 were not significantly different (t = -.68, two-tailed). These results are consistent with the representativeness heuristic: the source of the customer description did not matter, on average, to the subjects. The correlation between the subjects' responses to Parts 1 and 2 was
-.386. This was statistically significant ($p < .003$) and in the appropriate direction: the greater the probability of collection in full, the less the provision for uncollectible accounts should be.

Insert Figure 1

In this experiment, the data on both measures — probability assessments and decisions on the provision for uncollectibles — are consistent with the representativeness heuristic. To the extent that the customer description was provided by sources that are differentially reliable, as the auditing literature suggests, this finding is troublesome. Auditors may be insufficiently sensitive to the reliability of evidence.¹

Discussion. There are typically several means of obtaining evidence concerning the fairness of an account balance. In the case of evaluating the adequacy of the provision for uncollectibles, the means include the review of internal control,

¹An alternative explanation for these findings is that the independent credit agency and the client's credit manager were not perceived to be differentially reliable by the subjects. Another experiment reported in Joyce and Biddle [forthcoming(a)] rules out this rival hypothesis.
Experimental Condition

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Credit</td>
<td>Manager</td>
<td>Agency</td>
</tr>
<tr>
<td></td>
<td>.704 (.197)</td>
<td>.706 (.223)</td>
</tr>
<tr>
<td></td>
<td>25.3% (.244)</td>
<td>20.2% (.225)</td>
</tr>
</tbody>
</table>

Part 1 (Probability of Collection)

Part 2 (% Allowance for Uncollectibles)

Figure 1. Experiment 1 cell means as a function of question and experimental condition.

1Cell standard deviations are in parentheses.
analytical review, compliance tests, and substantive tests. The auditor uses his judgment to determine the appropriate combination of procedures, and the extensiveness and timing of their application so that his desired level of assurance is attained. If the auditor evaluates the audit evidence on the basis of their representativeness (and thus without regard for their reliability) the auditor may be unable to accurately assess the actual level of assurance attained through his audit tests. It is impossible to make an unconditional statement about the direction of the error. It will vary with the circumstances -- sometimes the actual level of assurance will exceed the desired level of assurance, and sometimes vice-versa. (The relationship between data representativeness and reliability in audit environments is not known.)

While the loss function for errors of overestimation of the level of assurance is probably quite different from the loss function for underestimates, both errors are costly. In the first case the extra evidence accumulated is not worth its cost to the auditor; in the second case an undesirable level of risk is assumed by the auditor.

We are presently conducting research on auditor sensitivity, to the differential reliability of internal vs. external documents (e.g., vendor's statements from the client's files vs. confirmations of accounts payable), and to the differential reliability of external documents (e.g., positive vs. negative confirmations).
IV. ANCHORING AND ADJUSTMENT

Judgments made in accordance with the anchoring and adjustment heuristic are hypothesized to result from the following process:

(1) An initial value or starting point (the anchor) which seems appropriate to the decision maker is selected. (Just how an initial value is chosen is not clear.)

(2) The anchor in (1) above is adjusted to take into account new information. This adjustment is typically in the appropriate direction but is insufficient in magnitude.

Take, for example, the problem of determining the magnitude and timing of substantive tests to perform in the audit of accounts receivable. A natural starting point would be the substantive tests as performed in the last engagement adjusted to allow for changes in circumstances between the two engagements — e.g., a weakening in the internal control system. If the auditor were employing the anchoring and adjustment heuristic as described above, he would modify his substantive tests in the normatively appropriate direction (i.e., he would increase the tests and/or change their timing to make them more diagnostic), but he would not go sufficiently far from the anchor. This could, of course, result in insufficient competent evidential matter being accumulated by the auditor to attain his desired level of assurance.

If internal control were strengthened, on the other hand, this heuristic would lead the auditor to reduce his substantive tests, but the reduction would not be as great as it should. Thus the auditor would be accumulating evidence not worth its cost.

Clearly, then, use of anchoring and adjustment in certain auditing decisions can lead to problems. Experiment 2 below was
designed to test for its existence in an audit context involving a change in internal control strength.

**Experiment 2**

The subjects in Condition A of Experiment 2 received the following problem:

You are conducting a routine year-end audit of a large closely-held tire wholesaler. Your firm has conducted its annual audit for the last three years. During that time no significant errors were discovered by the audit tests and unqualified opinions were issued. The management and employees of the client seem both competent and trustworthy. You are about to plan your substantive tests of the sales and collection cycle.

You review the client's system of internal control to identify the types of errors that could occur in the system and whether specific controls exist which would prevent or detect such errors. The first error you consider is "sales recorded for goods not shipped." Your review of the relevant controls for this error reveals the following:

1. There is adequate control over back orders and partial shipments.
2. There is adequate control of access to the shipping area.
3. Prenumbered shipping documents are used.
4. Sales invoices are matched to shipping documents.
5. Shipping is segregated from the billing function.
6. Regular shipping reports are prepared and reviewed monthly.
7. Overdue accounts receivable are investigated.
8. Unmatched sales invoices are independently reviewed and followed up.

Your compliance tests and observations of the system confirm the system is operating as just described.

1. Based on this information, indicate below the extensiveness of the substantive (detailed) tests you would perform in this engagement to test for the specific error "sales recorded for goods not shipped." (Circle one number.)

<table>
<thead>
<tr>
<th>Minimum Audit Tests</th>
<th>Most Extensive Audit Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4 5 6 7 8 9 10</td>
<td>10</td>
</tr>
</tbody>
</table>
2. Suppose that instead of the preceding system description, you had received one that was identical except for the following (assume your compliance tests and observations of the system confirmed this description as well):

(3) Prenumbered shipping documents are not used.
(5) The accounts receivable billing clerk also performs clerical duties for the shipping department.

Based on this different information, indicate below the extensiveness of the substantive (detailed) tests you would perform in this engagement to test for the specific error "sales recorded for goods not shipped." (Circle one number.)

<table>
<thead>
<tr>
<th>Minimum Audit Tests</th>
<th>Most Extensive Audit Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4 5 6 7 8 9  10</td>
<td></td>
</tr>
</tbody>
</table>

In part (1) of Condition A the subjects were asked to indicate the extensiveness of the substantive tests they would perform if the eight control features listed above part (1) were present. In part (2) they were asked to make the same judgment assuming control features 3 and 5 were absent. Thus in Condition A we observed how the auditors modified their substantive tests as internal control weakened.

The Condition B subjects were also asked to make two judgments concerning the extensiveness of the substantive tests they would perform. In this condition however, the part (1) judgment was in response to an internal control system where control features 3 and 5 were absent; in part (2) the judgment was in response to a system where features 3 and 5 were present. Thus in Condition B we observed how the auditors modified their substantive tests as internal control became stronger. Thus the Condition B subjects were asked to make the same judgments concerning
the same internal control systems as the subjects in Condition A, but in reverse order.

Subjects. The subjects participating in this experiment were 132 practicing auditors from the Chicago region of a Big Eight public accounting firm. The experiment was administered to the subjects on two occasions. On occasion 1, 55 auditors attending Firm 3's Chicago Region senior-level training school were tested, and on occasion 2, 77 auditors attending Firm 3's Chicago Region pre-senior or in-charge training school were tested. The experiments were conducted in a large conference room. The problem below was one of 10 problems on probabilistic judgment the auditors were asked to complete. The subjects were randomly assigned to either Condition A or Condition B, worked at their own pace, and completed the 10 experiments in 30-60 minutes.

Unlike the subjects who participated in Experiment 1, the Experiment 2 subjects were quite homogenous with respect to experience. Fifty-five percent had one year of audit experience and 40 percent had two years of audit experience.

Results. The cell means for Experiment 2 are presented in Figure 2. Figure 3 is the sources table for the 2 x 2 repeated measures ANOVA performed on the data.

Insert Figures 2 and 3
<table>
<thead>
<tr>
<th>Order Condition*</th>
<th>A (stronger first)</th>
<th>B (weaker first)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Stronger</td>
<td>Weaker</td>
</tr>
<tr>
<td></td>
<td>2.78</td>
<td>6.29</td>
</tr>
<tr>
<td></td>
<td>3.19</td>
<td>5.44</td>
</tr>
<tr>
<td></td>
<td>2.99</td>
<td>5.87</td>
</tr>
</tbody>
</table>

*Factor manipulated between subjects.

**Factor manipulated within subjects.

Figure 2. Experiment 2 cell means as a function of internal control strength and order.
<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between subjects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Order</td>
<td>131</td>
<td>3.28</td>
<td>0.50</td>
<td>.4792</td>
</tr>
<tr>
<td>Subject within groups</td>
<td>130</td>
<td>6.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within subjects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internal control strength</td>
<td>132</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Order x internal control strength</td>
<td>1</td>
<td>545.72</td>
<td>331.74</td>
<td>.0001</td>
</tr>
<tr>
<td>Internal control strength x subject</td>
<td>130</td>
<td>1.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>within groups</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3. ANOVA sources table for Experiment 2.
A significant main effect was observed for the internal control manipulation ($F = 331.74, p < .0001$). On average, the subjects indicated that more extensive audit tests should be performed when internal control was weaker than when it was stronger. This finding is consistent with both normative principles and use of the anchoring and adjustment heuristic.

No significant main effect was observed for order. This simply means that the mean responses across internal control conditions did not differ significantly between those subjects getting the stronger internal control system first and those getting the weaker internal control system first. This finding provides evidence neither for nor against the use of anchoring and adjustment.

Finally, a significant order x internal control strength interaction was observed ($F = 15.98, p < .0002$). The interpretation of this interaction is particularly interesting: the effect of a particular level of internal control is contingent on whether this level of internal control was provided first or last to the subjects. An order effect is obviously at work here and shown clearly by Figure 2. The subjects in Condition B start with weaker internal control and indicate less extensive work than the Condition A subjects who finish with the weaker internal control (5.44 versus 6.29). This may represent radical rather than insufficient adjustment by the Condition A subjects who go from stronger to weaker internal control. Note also that the subjects in Condition A start with stronger internal control and indicate less
extensive work than the Condition B subjects who finish with the stronger internal control (2.78 versus 3.19). Here we may have a case of insufficient adjustment on the part of the Condition B subjects as they go from weaker to stronger internal control.

Clearly, the anchoring and adjustment heuristic cannot account for all that occurred in this experiment. It can explain the possible insufficient adjustment but not the possible radical adjustment. Is there something that can account for both? Note that a weaker system looks weaker than it perhaps should after a stronger system. Note also that a stronger system perhaps looks weaker than it should after a weaker system. It appears that the auditors may be doing something consistent — understating the strength of the second system.

Some implications of this finding are clear. In continuing engagements it is common to take last year's audit program and modify it for changes that have occurred in the client's system in the current period. The results here suggest that the adjustment made may lead to too much work being performed regardless of whether the system has improved or worsened since the last audit.

V. AVAILABILITY

In some situations people estimate frequencies or probabilities of events on the basis of how easily they can be recalled from their memories or generated by their imaginations. An auditor might estimate the probability of being sued if his client goes bankrupt by remembering instances where other auditors have been sued following the bankruptcies of their clients. He might
estimate the probability of a material error in an account by imagining ways in which a material error could occur in spite of the client's internal control system.

Obviously, making probability assessments on the basis of cognitive availability is often quite useful. In some situations it is the only feasible basis for making predictions about the likelihood of uncertain events. In general, events which happen frequently are more easily recalled or imagined than events which occur only rarely.

If the ease of recall or imaginability of events were affected only by their frequency of occurrence, availability would be a reliable heuristic for estimating probability. Ease of recall, however, is also affected by factors such as salience and recency. Seeing a person shot to death on the street will have a much greater impact on a person's subjective probability of such a crime than hearing about it on the radio. Events which have happened yesterday are easier to remember than events which happened a year ago. Because availability is affected by these factors unrelated to frequency, its use can lead to systematic errors in assessing probabilities or frequencies. In Experiment 3 below we attempted to manipulate the availability of the ways in which standard audit procedures could detect a particular employee fraud.

Experiment 3

The subjects in Condition A of this experiment were asked
to work the following problem:

You are examining the sales invoice documents of a medium-sized mail-order retailer (about 5,000 orders per month). The client's personnel manager has informed you that he is concerned about the integrity of several of his employees. He is reluctant to tell you their names but has indicated that at least one is in the order receipt department where customer orders are received and two others are in the warehouse control department where customer orders are assembled and shipped. In the order receipt department some of the clerks are assigned to open customer orders, record the order number and amount of payment received and forward a copy of the order to the warehouse control department. The customer checks are sent to the accounting department where bank deposits are prepared. The original copy of the customer order is filed in the order receipt department until a notification list from the accounting department is received indicating the customer's check has cleared. At this point the customer order is approved by being stamped PAID and the warehouse control department is notified that shipment can be made. The original and warehouse control copies (which are stamped SHIPPED) are forwarded to the accounting department where they are filed together in a closed customer order file (though separate from photocopies of the checks). A log is kept of all shipments and suspense files are kept in the order receipts department and the warehouse control department for those orders which have not yet been approved.
In light of the personnel manager's comments you have reviewed the system of internal control and you are generally satisfied with the document flow and segregation of duties. You are still concerned, however, about the possibility of shipments being made on fictitious invoices which have been approved by a dishonest clerk in the order receipt department. There are, of course, several ways this type of employee fraud could be detected using standard audit procedures. For example, there should be a photocopied check and corresponding bank deposit entry for each customer order on file, each customer order should be recorded in the order receipts log and each order should be listed on the check clearance notification list. All customer orders on file should be marked PAID and all shipment copies should be accompanied by a PAID customer order copy. The entries on the shipment log should correspond to the check clearance list and the suspense files in the order receipt department should correspond with those orders in the warehouse control department's suspense files.

Suppose that the personnel manager had not informed you of his suspicions concerning the integrity of the employees. In your estimation, what is the probability that standard audit tests would have detected this type of employee fraud if not more than 50 (out of approximately 5,000) fictitious shipments are being made each month? (Indicate this probability by circling the number closest to it.)

0 .05 .10 .15 .20 .25 .30 .35 .40 .45 .50 or greater
The subjects in Condition B received a problem identical to this except that no examples of how standard audit procedures might detect the fraud are provided. By providing these examples for the subjects in Condition A, we hoped to increase the availability of the ways in which the employee fraud could be discovered. If the availability manipulation were successful, we would observe the Condition A subjects estimating a higher probability of detection than the subjects in condition B.

**Subjects.** The same 50 subjects who participated in Experiment 1 participated in this experiment. They were randomly assigned to the experimental conditions.

**Results.** The cell means for Experiment 3 are reported in Figure 4. While the condition A cell mean is higher than the cell mean for Condition B, there is no statistically significant difference ($t = .37, p = .714$). The data do not support the availability hypothesis.

Insert figure 4

After examining the data and reexamining Experiment 3, we feel that our failure to observe behavior consistent with availability may be due, at least in part, to problems with our experiment. Our experimental manipulation was designed to make the ways in which standard audit procedures could detect the employee fraud more available to the Condition A subjects than the Condition B subjects. The manipulation may not have worked for at least two
Experimental Condition

<table>
<thead>
<tr>
<th>High Availability</th>
<th>Low Availability</th>
</tr>
</thead>
<tbody>
<tr>
<td>.196 (.164)</td>
<td>.178 (.181)</td>
</tr>
</tbody>
</table>

Figure 4. Probability of detection as a function of availability.

1Cell standard deviations are in parenthesis.
reasons. First, it may have been quite easy for the Condition B subjects to imagine or remember (from previous experience) what we provided the Condition A subjects. Second, this particular example of employee fraud may have been viewed as so hard to detect (with not more than 1% of the shipments fictitious) that differential availability made no difference. At this point we have doubts we have conducted a good test of availability. We intend to try again.

VI. GENERAL DISCUSSION

This paper has reviewed the results of three experiments and has reported rather mixed findings. In Experiment 1, auditors appeared to ignore the reliability of evidence and behave consistent with the representativeness heuristic. A possible implication of this finding was failure of the auditors to attain the desired level of assurance.

Experiment 2 was a test for anchoring and adjustment. The results obtained were not consistent with this heuristic, nor consistent with normative principles as auditors' recommendations concerning the extensiveness of substantive tests were subject to order effects. Apparently some as yet unidentified heuristic was at work.

Experiment 3 did not produce behavior consistent with the availability heuristic, but our ex post reservations concerning the definitiveness of the experiment have led us to temporarily suspend judgment.
As we mentioned above, the primary reason auditing researchers are concerned with heuristics is that such study might improve the quality of audit decision making. The type of research discussed in this paper is merely a first step in this direction. Before decision making can be improved, it is necessary to know in what areas it is deficient. Heuristic use per se does not provide conclusive evidence of deficiency. It is necessary that such use lead to systematic, costly decision errors. As we mentioned in the introduction, heuristics work quite well in many tasks. Thorngate (in press) presents evidence from a simulation showing how heuristics lead to optimal responses across a variety of tasks. Even where heuristics lead to systematic, costly errors, the cost of adopting an alternative procedure which eliminates the error might outweigh the incremental benefit.

In spite of these problems we feel that additional research wherein heuristics suggest potential deficiencies in audit decision making is advisable. Research on improving the quality of intuitive judgments where they are found wanting has barely begun. (See Kahneman and Tversky, in press; Kinney and Uecker, 1980 for examples of judgment aids.) Efforts in this area are likely to be impeded until the psychological theory of decision making is better formulated. Such a formulation must include a concern for task structure, the cognitive representation of the task, and the information processing capabilities of the organism (Einhorn and Hogarth, in press).
An example of the complexity of the interaction between
task and cognitive representation is provided by Tversky and
Kahneman (1980).

Problem 1. Imagine that the U.S. is preparing for the outbreak
of an unusual Asian disease, which is expected to kill 600
people. Two alternative programs to combat the disease have
been proposed. Assume that the consequences of the programs
are as follows:

If Program A is adopted, 200 people will be saved.

If Program B is adopted, there is a 1/3 probability that 600
people will be saved, and a 2/3 probability that no people will
be saved.

Which of the programs do you favor?

Seventy-six percent of the subjects chose Program A. A
second group of subjects received an equivalent problem, but with
a different formulation of the two programs:

Problem 2.

If Program C is adopted 400 people will die.

If Program D is adopted there is a 1/3 probability that nobody
will die, and a 2/3 probability that 600 people will die.

Eighty-seven percent of the subjects chose program D. Note
that Programs A and C are identical as are Programs B and D.
Merely changing the descriptions of the outcomes from lives saved
to lives lost produced a pronounced change in preference.

This sensitivity of cognitive representations of tasks to
normatively irrelevant aspects of tasks might have important
implications for auditor education. Firms may wish to structure
audit decision problems in such a way that leads auditors to adopt
cognitive representations of the problem conducive to decision
making in conformance with normative principles. We are currently
investigating this possibility.

Clearly, much work remains to be done in this area before
significant payoffs will accrue to auditors. As Curt Gowdy would
say, research on heuristics and biases "has most of its future
ahead of it." We feel this future presents auditing researchers
with the opportunity to contribute significantly to the quality of
audit decision making.
BIBLIOGRAPHY


Miller, G. The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychological Review 63 (March 1956): 81-97.


Heuristics and Biases: Some Implications for Probabilistic Inference in Auditing: A Discussion

by
Gary L. Holstrom
Deloitte, Haskins & Sells
I am grateful for the opportunity to review the paper by Biddle & Joyce because I have found their work to be helpful to me in analyzing the role of heuristics in audit judgment within Deloitte Haskins & Sells. This year, I wrote a paper on audit judgment for our firm's Audit SCOPE Program for auditing professors that included a discussion of the positive contribution made by the 23 experiments conducted by these researchers. Although the specific experiments reported in the paper presented here by Biddle & Joyce yielded somewhat mixed and disappointing results, they should be viewed in the context of the positive contribution made by their experimental program as a whole.

The paper by Biddle & Joyce reviews the results of three experiments, each designed to test whether auditors' judgments are consistent with the use of one of three identified heuristics—representativeness, anchoring and adjustment, and availability. Unfortunately, none of the three experiments provided conclusive evidence of the presence or absence of any of the identified heuristics. The authors' own discussion comments indicate that the results were inconclusive with respect to Experiment 2 (concerning anchoring and adjustment) and Experiment 3 (concerning availability). Furthermore, a closer look at the results of Experiment 1 (concerning representativeness) reveals that this experiment also failed to adequately test for the presence or absence of the heuristic. I will comment on the results of each of these experiments.
Experiment 1 - Representativeness

The intended purpose of Experiment 1 was to test one aspect of the representativeness heuristic - the tendency to ignore the reliability of data supplied in a description of a client's past-due customer when the description appeared to be representative of customers who will pay their accounts. Actually, the experiment was a test of the extent to which auditors' estimates of the probability of collection of an account receivable were sensitive to presumed differences in their perceptions of information reliability.

Conceptually, the experimentally manipulated variable was the perceived reliability of the data in the description. A basic tenet of experimental design is that if the experimentally manipulated (predictor) variable is not significantly different between experimental groups, one could hardly expect to find a significant difference in the criterion variable. The central question is whether the intended experimental variable, perceived reliability, was adequately differentiated between groups.

The experiment utilized a "between-subjects" design in which participating auditors were divided into two experimental groups that were provided a page of information that differed only with respect to one clause of one sentence, concerning the source of the customer description. For one group the instructions contained a sentence stating: "You investigate the customer further and get the following description from the client's credit manager." For the other group the instructions contained a sentence stating:
"You investigate the customer further and get the following description from an independent credit agency which has provided you with reliable and up-to-date information in the past."

This raises the crucial research question of whether this difference in a single clause in the instructions was sufficient to result in a significant difference in the perceived reliability of the data in the description.

If both sets of instructions were to be presented in sequence to the same subject, the differentiating sentence would likely be identified and attention would likely be given to the question of whether different estimates of data reliability were warranted. However, if only one set of instructions is presented to a subject, the development of a clear perception of the reliability of the data is confused considerably.

For example, in the "Credit Manager" case, the auditor is told: "You investigate the customer further and get the following description from the clients' credit manager." The subject may imply from this statement that the auditor has conducted a further investigation of the customer in addition to obtaining the description from the credit manager. Even if the subject gives attention to the question of the reliability of the data supplied by the credit manager, several questions remain:

1. If the description is "from the client's credit manager", where did the data in this description originate:
   (a) with an independent credit agency?
   (b) with the client's discussions with the customer's management?
   (c) with another business or social source?
2. If the auditor has investigated the customer further, as stated in the instructions, would the auditor have been likely to obtain evidence to corroborate the client's description.

3. Has the credit manager provided reliable & up-to-date information in the past?

All of these factors raise doubts concerning the adequacy of the experimental controls and concerning whether the instructions for the two groups were sufficiently different to yield statistically significant differences in perceived reliability. The experiment results shown in Figure 1 indicate that the differences between the groups—though extremely slight in one of the two cases—is in the direction that is consistent with normative theory; the differences are just not statistically significant.

Since the failure to detect significant differences may well be attributed to the lack of power of the "between-subjects" design, it may potentially be overcome by using a "within-subjects" design. The authors mention this alternative explanation of the research findings in Footnote 1 on Page 9, but indicate that the alternative hypothesis is ruled out by the results of another experiment conducted by the researchers and reported elsewhere (see Joyce & Biddle, "Are Auditors' Judgments Sufficiently Regressive?" University of Chicago working paper; October, 1979; forthcoming in the Journal of Accounting Research). I disagree with the authors on this point. A close look at the other experiment tends to substantiate the rival hypothesis, not rule
it out. The other experiment (Experiment 3B in the authors forthcoming article on judgment regressiveness) utilized a within-subjects design. Subjects were presented the customer description, told the source of the description (either credit manager or credit agency) and asked to indicate the probability of collection and the appropriate percentage to be established as an Allowance for Doubtful Accounts. Then they were asked to judge the probability of collection and appropriate allowance percentage if the description had been provided by the other source. The sequence in which the sources of the data were presented was reversed for one half the subjects so that the effect of sequence was controlled and tested.

Results of this other experiment clearly showed that when the source of data was made salient in the more powerful "within-subjects" design, the difference in source reliability resulted in statistically significant differences in auditor judgments in the normatively appropriate direction. Estimates of probability of collection based on credit manager data were lower than those based upon an independent credit agency. Likewise, auditor judgments concerning the appropriate percentage to include in the allowance account were higher when based upon the data supplied by the credit manager.

In my opinion, the most reasonable conclusion from these two experiments is that the failure to obtain significantly different results in Experiment 1 is most likely attributable to the insufficient manipulation of the experimental variable in the less powerful "between-subjects" design. The results cannot justifiably
be regarded as empirical evidence of the presence of the representativeness heuristic in this case. Further evidence to support this conclusion is provided in the other experiment mentioned in the authors' Footnote 1. When the distinction between sources of data was clarified in the other experiment, auditor judgments were significantly modified in the normatively appropriate direction. These experiments clearly did not demonstrate the presence of the representativeness heuristic.

Experiment 2 - Anchoring and Adjustment

The second experiment was intended to test whether auditor subjects tended to use the anchoring and adjustment heuristic. Results of this experiment also were inconclusive, leading the authors to state:

The results obtained were not consistent with this heuristic, nor consistent with normative principles as auditors' recommendations concerning the extensiveness of substantive tests were subject to order effects. Apparently some as yet unidentified heuristic was at work. (p. 21)

The results of this experiment were so mixed that the authors found themselves resorting to what seemed like an ex post fishing expedition for plausible explanations - a situation that is a potentially fruitful basis for developing ideas for future research but a tenuous and even dangerous basis for building theory and for developing implications for practice. In discussing results that merely show a greater magnitude of
adjustment in one order condition (strong first) than in the other (weak first), the authors describe the former as a "possible radical adjustment" and the latter as a "possible insufficient adjustment." Actually, in neither condition do the results indicate whether the adjustment was radical, insufficient or exactly right in a normative sense. Normatively, it is possible that both adjustments may have been insufficient or that both may have been radical. The point is that the discussion of what was possibly radical or possibly insufficient was all relative to the composite judgment of the other group, a measure that was not at all normative.

The problem with the researchers' ex post hypothesizing in this case lies in their conclusion: "It appears that the auditors may be doing something consistent - understating the strength of the second system." (p. 183) The results provide no more support for this hypothesis than for an alternative hypothesis that auditors may be overstating the strength of the first system. The experiment provides no basis for concluding that the first judgment in the sequence is better than the second, or vice versa.

In their ex post search for explanations of the inconclusive results, Biddle & Joyce unjustifiably went beyond the point of suggesting possible avenues for future research. Instead, they suggest that their ex post explanations have clear practical implications, as follows:

Some implications of this finding are clear. In continuing engagements it is common to take last year's audit program and modify it for
changes that have occurred in the client's system in the current period. The results here suggest that the adjustment made may lead to too much work being performed regardless of whether the system has improved or worsened since the last audit. (p. 16)

In my opinion this statement is not supported by the data, which merely show that a given internal control condition was judged, on the average, to be more favorable when it was presented second in the sequence than when it was presented first. The data provide just as much support for the hypothesis that too little work may be performed in the first year as they do for the hypothesis that too much work may be performed in the subsequent year. The point is that, based on the data alone, neither implication has been substantiated.

A somewhat different approach—and to me a more promising one—for testing for the presence of the anchoring and adjustment heuristic would be to supply the same internal control description to both groups. One group could be told that other auditors had rated the controls as being relatively strong (for example a 2.0 on the "extent of tests" scale) and the other group could be told that other auditors had rated the controls as being relatively weak (for example a 7.0 on the "extent of tests" scale). This would provide a clearer test of the effect of supplying different audit-related anchors.

Experiment 3 - Availability

I have little to say about the third experiment because I share
the researchers doubts about whether they have conducted a good test of availability. The data neither support nor refute the availability hypothesis because it was possible--indeed, quite likely--that the experimental group which did not receive the descriptive examples may, nevertheless, also have had an awareness (availability of the concept) of how audit procedures could detect the fraud.

Such awareness of the ways in which audit procedures could detect fraud was likely available to both groups. Either such an understanding was available to both groups, or else there has been a drastic failure in the auditing education process, the validity of the CPA exam, and the training programs of the participating firms. Although I grant that consideration of the possibility of such failures should not necessarily be ruled out as a topic for empirical research, I maintain that heuristic experiments should not be built upon an implicit premise that such total failures currently exist.

In summary, I believe the authors are to be commended for the contribution of their program of 23 experiments, taken as a whole. However, none of the three experiments presented in this paper provided an adequate test of the presence or absence of the intended heuristic.
Heuristics and Biases: Some Implications for Probabilistic Inference in Auditing: A Discussion

by

Barry L. Lewis

University of Pittsburgh
Biddle and Joyce have given us a taste of their extensive research efforts into the judgmental heuristics and associated biases of professional auditors. The relatively recent attempt to structure audit problems within a normative decision theoretic framework has encouraged a growing volume of research into how auditors process probabilistic information. An example of this growth can be seen in successive reviews of this literature (Libby and Lewis, 1977; 1980) in which the number of studies classified under probabilistic judgment increased from seven to twenty-two. Studies dealing specifically with heuristics and biases increased from one to eight. These latter studies have evolved from simple replications of Tversky and Kahneman (1974) experiments to increasingly complex extensions in audit contexts. But it is apparent that we have not yet scratched the surface of understanding how auditors make probabilistic judgments.

The general motivation for this type of research is to improve decision making by developing decision aids to overcome the systematic biases which may result from generalized heuristic processing of information. This involves testing for the presence of specific heuristics; comparing judgments with normative solutions and measuring the cost of systematic deviations; developing and testing decision aids to eliminate biases; and implementing the decision aids (if they are cost-effective). The reasonableness of this research approach rests on certain assumptions. First, we assume that the normative model is an appropriate benchmark. Second, we assume that we can measure the costs of non-normative behavior. Third, we assume that we can obtain valid experimental evidence of non-normative behavior.

While specifying the cost of deviation from normative judgment may be a formidable measurement problem, most of us would probably agree that it does not constitute a conceptual roadblock to applied decision research. Likewise, many of us would probably accept Bayesian revision and expected utility
maximization as appropriate models for probabilistic processing and decision making. On the other hand, I must point to cogent arguments by Einhorn & Hogarth (1981) against the unquestioned acceptance of normative models. These arguments relate to how well the normative models fit the task environment. They also point out the dilemma of competing normative models. The third assumption — that of valid experimental results — is achievable at least in theory. The remainder of this discussion will focus on the validity, in practice, of the experiments of Biddle and Joyce.

In the first experiment, subjects are to judge the probability of collection of an account receivable given the base rate of uncollectibles and a customer description rendered by either an external credit agency (Condition A) or the internal credit manager (Condition B). Results showed that the mean probability estimates between conditions were not significantly different. Can we infer from these results that auditors do not consider source reliability in audit judgments or can we generate other plausible explanations for these results? Failure of the independent variable manipulation would be one such explanation. Auditors may not think there is a difference in reliability of the two sources, for example. In another version of this experiment, reported elsewhere (Joyce & Biddle, forthcoming), the authors used a within subjects design and found evidence of differential perceptions of source reliability. This would seem to indicate that, in the absence of the "red flag" aspects of a within subjects design, auditors are insensitive to source reliability. But are the subjects seeing the same situation that the experimenters have intended? In an actual audit situation, the question of source reliability may be more relevant in evidence collection than in evidence evaluation. Having been presented with a fait accompli, the subjects may ascribe reliability to the experimenter rather than to the titular source of the description. A possible solution to this dilemma is to present a richer situation in which subjects search for evidence. By design, some subjects depend far more...
credit agency's report would be able to receive that information. The remaining subjects would be told that the credit agency has no data on the customer and would be given the credit manager's report. But all subjects would be given the opportunity to search for the information they need. This would avoid the artificial red flag of the within-subjects design and would avoid the lack of realism in the between-subjects design.

In the second experiment, subjects make successive extent-of-audit decisions given two descriptions of internal control (strong and weak). The between-subjects manipulation is the order in which the two descriptions are presented. The authors predicted the classic adjustment bias as an insufficient move from the first judgment - the anchor. This classic bias would mean that auditors would tend to over-audit when moving from a weak to a strong system and would tend to under-audit when moving from a strong to weak system. Experimental results indicate that auditors would always tend to over-audit when the system changes, a condition the authors attribute to conservatism (although not in the Bayesian sense). This experiment is a valid test of the anchor and adjustment heuristic only if the change information is normatively irrelevant. If we assume that an extent-of-audit judgment is a function of the probability of error, we must ask ourselves whether the probability of error given a description of weak controls is equal to the probability of error given that the control system has deteriorated from a strong to a weak state. I do not think they are equal. If they are not equal, these experimental results are uninterpretable vis-a-vis a normative model.

In the third experiment, subjects are required to estimate the probability of detection of employee fraud through standard audit procedures. To increase the "availability" of likely procedures, one group of subjects are given examples of how the fraud might be detected. There was no difference in detection estimates between the two groups. In this experiment, the authors provide their own explanation for the failure to detect evidence of the availability heuristic.
Significantly, it is the same explanation that I provided for the first two experiments — failure of the experimental manipulation. In each of the three experiments, we really have no idea of what the subjects are doing, why they are doing it or how they define and delineate the task. Unfortunately, when we get results consistent with our hypotheses, we tend to infer a process consistent with our hypotheses. In many respects, we may have learned more from experiment three than from the first two because we were forced to look beyond the responses of the subjects.

I have suggested three stumbling blocks for this type of applied judgment research — the appropriateness of the normative model, the ability to measure the cost of non-normative behavior, and the validity of experimental results. While I have discussed only experimental validity issues, I do not want to minimize the importance of the other issues. Nor do I want to denigrate the research of Biddle & Joyce. Their work is a necessary and useful step toward understanding how auditors make probabilistic judgments. But in research which has strong policy implications, it is quite important that we be confident that our experimental results reflect what is actually happening or what would actually happen in a field environment. Biddle & Joyce have begun to extend the work of Tversky & Kahneman beyond context—free riddles to meaningful professional problems. The next steps, however, must be toward more realistic situations that can involve the subjects in a more active way. And in this more complex environment we must be able to trace the activity of the subject in his definition of the task, his information search, his hypotheses and his decisions. As a model for this type of high fidelity research, let me recommend the experiments in medical problem solving by Elstein, Shulman & Sprafka (1978).
References


MODELING AUDITOR JUDGMENT
BASED ON TWO METHODS OF AUDITOR
RATIONALE DOCUMENTATION

by

Theodore J. Mock

and

Paul R. Watkins

Working Paper No. 12
Center for Accounting Research
School of Accounting
University of Southern California

Prepared for presentation at the fourth biennial Symposium on Auditing Research, University of Illinois, November 13 and 14, 1980. Comments are welcomed.
An auditor renders an opinion upon the fairness of presentation of results of operations for a given period of time and upon the fairness of presentation of the financial position at the end of that given period of time. To develop an opinion as to fairness of presentation, the auditor must gather and evaluate many different types of information, both financial and non-financial. This gathering and evaluation activity is known as the audit process.

Central to the audit process is the auditor's judgment process. This judgment process involves evaluations of available audit evidence, professional standards relating to audit evidence, and the auditor's personal criteria as to fairness of presentation. An important result of the auditor's judgment process is the audit program upon which the actual audit will be based, and which will ultimately influence the "fairness of presentation" decision.

To encourage a consistent approach to this judgment process, most auditors and audit teams have developed some type of sequential decision system in which each judgment builds upon the results of previously performed procedures and judgments. One of the key elements in this sequential process is the identification, documentation, and evaluation of internal accounting controls.

Mock and Turner (1980) have presented a comprehensive description of the internal control evaluation process and a number of experimental results which provide insights into the internal control evaluation aspect of auditor judgment. Of particular interest to this paper is their findings with respect to unexplained variability in auditors' decisions. Significant variability measured in terms of both ranges and coefficients...
of variation were observed in the five related Mock and Turner experiments. But, more importantly, much of the observed variability was not attributable to experimental variables or interpersonal differences. Analysis of variance showed that the two experimentally controlled variables (compliance test results and auditor guidance) accounted for only 10 to 23.6 percent of the observed variance in the auditors' four scope decisions. These and other Mock and Turner results imply that additional research is needed regarding the ways in which auditors formulate a judgment, that is, the way in which they weight various informational, organizational, and behavioral cues and combine them in order to reach a decision.

Another important issue raised in the Mock and Turner study is the adequacy of auditor judgment documentation. The adequacy issue is addressed in this paper by attempting to use two forms of auditor rationale documentation to build improved models of auditor judgments. Hopefully, models that explain more variance and exhibit predictive validity can be derived.

**Alternative Approaches to the Modeling of Judgment and Decision**

Interest in the modeling of decisions has been evident in accounting research for many years (Bruns, 1962; Mock, 1969) although only recently have some of the more sophisticated behavioral science techniques been applied to auditing (Ashton, 1974; Joyce, 1976; Biggs and Mock, 1980). Audit judgment research has examined auditor decision making from several perspectives, although the primary methodology used is experimentation. This type of research may be classified as **synthetic a posteriori** (Wartofsky, 1968).
Synthetic A Posteriori Models

Models which are derived through observation (i.e. a posteriori) are of most interest in auditing because complexity makes the derivation of analytical models difficult. Within this class, the models may be idiographic (individual specific) or nomothetic (based on common features shared by individuals). Little idiographic judgment modeling research has been published in accounting and auditing. An exception is Biggs (1979) who has modeled judgments for several investment analysts.

As part of the Mock and Turner study, a protocol study was conducted (see Biggs and Mock (1980a, 1980b)) which derived individual specific information search and choice models for several auditors. Results indicated that two approaches to information search were evident, but much of the variability in the auditor's judgments observed in the main, experimental portion of the Mock and Turner study remains unexplained.

The main purpose of this paper is to present the results of a synthetic a posteriori modeling approach to auditor judgments. In general, this approach involves imputing models using multivariate statistical techniques applied to repeated measures of auditor judgments (as dependent variables) and various input cues, demographics or other possible independent variables.

Two sets of data with which to build these models are available (see Table 1). The first set is the content-analyzed rationale memos many auditors prepared for the Mock and Turner study. These memos were prepared for each of four audit sample size decisions that were made. As these memos tended to be prepared after the decisions were reached, they are labeled retrospective documentation in this paper. Details
concerning the content analysis procedures, including reliability and validity considerations, are discussed in Mock and Turner (1980).

The second set of data labeled concurrent documentation consists of a series of judgments a subset of the auditors were required to make as they evaluated the same four audit procedures. These judgments were elicited in a "structured documentation form" (a questionnaire).

DATA GATHERING AND TASKS

The dependent variable data sets consist of sample size judgments of audit seniors and supervisors. The retrospective data set reflects sample size judgments of 73 auditors and the concurrent data set reflects the judgments of 32 auditors. Each auditor was presented a case containing information on improvement in internal accounting controls. They were asked to adjust, as they considered necessary, the planned sample size for four specific auditing procedures in an audit program. The experimental case study was based on a portion of a commercial entity's revenue cycle and presented the subjects with nearly all the documentation normally available during an audit. The documentation included consisted of:

1. The prior year's completed audit program and the current year's partially completed program;
2. A memo summarizing audit planning (economic, organizational, management, general control environment, and other data);
3. Accounting system documentation (flowcharts);
4. "Bridging work papers" which related audit objectives, system controls, compliance tests, and subsequent substantive procedures;
5. Results of interim compliance tests, which provided evidence of improvements in specific internal accounting controls;

6. Miscellaneous data including interim financial information and results of the prior year's confirmation tests.

Subjects were instructed that they were replacing an audit senior who had resigned in mid-audit to take a position in industry, and that they should perform the following tasks:

1. Evaluate the planned audit program for four procedures with respect to extent of testing.
   Procedure E-5: Packing slip - invoice comparison
   Procedure E-6: Invoice pricing tests
   Procedure E-9: Posting tests
   Procedure E-10: Confirmation of accounts receivable

2. Prepare rationale documentation for each of the four audit procedures. Two types of documentation were prepared. Retrospective documentation consisted of open-ended rationale memos prepared after the subjects had arrived at their sample size recommendations. Concurrent documentation consisted of filling out detailed questionnaires as each subject was gathering data and reaching a sample size decision.

The essential difference between the two documentation tasks was one of timing. That is, in the concurrent documentation task, auditors were asked to consider various factors and to indicate judgments concerning these factors in reaching their explicit sample size judgments. On the other hand, the retrospective documentation task required the auditors to
recall, after the fact, which factors had influenced their sample size recommendations.

Research Questions

Based on the data from the above experimental setting, and the desire to consider the issues of modeling auditor judgment and appraising documentation adequacy, the following research questions were developed:

1. What is the nature of the decision (judgment) models underlying sample size judgments based on concurrent and retrospective documentation techniques?

2. What is the relative importance of the various factors in the models?

3. Are the derived cue weights assigned to the various factors stable across audit procedures?

4. Do the descriptive, derived models have predictive validity?

5. What are the effects of individual auditor differences on the model parameters?

6. Are there significant differences between the concurrent documentation and the retrospective documentation models? If so, what is the basis of similarities/dissimilarities in the two models?

The above research questions are concerned with aggregate or group models of auditor judgments, and thus are nomothetic in methodological terms.
DATA ANALYSIS METHODS AND RESULTS

The research questions of the previous section reflect several important issues underlying the modeling approach. These issues, broadly stated, are categorized as follows:

1. Model search and development -- research questions 1, 2, 3.
2. Model interpretability -- research questions 2, 3.
3. Model testing -- predictive ability -- research question 4.
4. Model representativeness -- nomothetic versus idiographic considerations -- research question 5.
5. Documentation timing effects -- research question 6.

Each of these issues is now discussed with a description of the analyses utilized in their assessment, and the ensuing results. A discussion of the results constitutes the final section of this paper.

Model Search and Development

The starting point for model development is an assessment of the nature of the underlying data. As previously stated, the two documentation approaches required the auditors to indicate their use of, reliance on, and/or judgments concerning certain factors in reaching a sample size decision.

Factors included in the content analysis dictionary (see Table 1) for the retrospective documentation task are: audit test objectives, general and specific control references, cost, risk, other audit evidence, nature and timing of the test, and nature of the test: compliance, substantive, dual. For the concurrent documentation task, the questionnaire elicited judgments are: nature of the audit test: compliance, substantive, dual;
materiality, degree of reliance, and sample size range indicated by maximum sample size and minimum sample size. A more detailed description of these factors for each of the two tasks is presented in Table 1.

The majority of the data, reflecting reference to these factors, is binary coded with a "1" indicating reference to or reliance on a factor in the documentation task and a "0" indicating no reference. For the concurrent documentation task, minimum and maximum sample sizes were ratio scaled.

In addition to the reference factors, demographic information was gathered and evaluated for the retrospective documentation task. This information is included in this study in an attempt to see if additional variance can be "explained" for the judgment models. The demographic factors are: actual number of years of audit experience (ratio scaled by actual years), specialist status (coded 0 if not a specialist, 1 if computer or statistical specialist), special firm computer audit training course (coded 0 if no, 1 if yes), local office internal control training course (coded 0 if no, 1 if yes), local office statistical training course (coded 0 if no, 1 if yes), commercial experience (coded 0 if none, 3 if extensive), number of audit level training courses (coded in intervals: 0 if lowest and 4 if highest), and client mix (coded 1 if primarily small, 3 if primarily large).

One additional information factor was included and assessed in the model development. This was binary coded data reflecting the experimental treatments (0 if internal controls were "fair", 1 if "strong") of the Mock-Turner (1980) study. This factor was included since between 10 and 24 percent of the sample size judgment variance has been previously explained by this experimental treatment.
Table 1
FACTORs USED TO ANALYZE RETROSPECTIVE DOCUMENTATION

Reference 1 - Audit Test Objective

Examples of possible objectives referenced:

a. Validity of recorded transactions
b. Proper authorization of transactions (balance)
c. Assignment of proper initial economic value for recording purposes
d. Accurate recording of transactions
e. Proper valuation of transactions to reflect current economic value

Reference 2 - Reference to General or Specific Controls

Reference to Evaluation of General Controls (e.g., "controls are strong")
Reference to Specific Controls e.g.

Prenumbered sales invoices are:

a. Prepared for all sales
b. Issued sequentially
c. Numerically accounted for

Reference 3 - Cost/Benefit Factors

Some statement as to cost/benefit factors of the procedure or evidence generated (e.g., would combine with other step, procedure gives limited results, step is justified, enables us to limit confirmation effort, time could be better used, does not serve a useful purpose)

Reference 4 - Audit Risk in Account, Item Being Audited

Some mention of audit risk (e.g., possibility of error, understatement, overstatement, and/or materiality, goods billed do not correspond to goods shipped, our "exposure", accuracy, missing invoices, shipments with no corresponding billing)

Reference 5 - Reference to Other Audit Evidence Relied Upon

(e.g., confirmation replies, analytical review of cost of goods sold, substantive tests in the previous year)
Reference 6 - Nature and Timing of the Audit Test

Some reference to the need for changing the nature and/or timing of the test.

Reference 7, 8 - Nature of the Test

Reference to nature of the test (e.g., substantive, compliance, or dual).

Reference 9 - Compliance Test Results

Some mention of the results of completed compliance tests (e.g., no exceptions were noted in audit procedure E-7)

VARIABLES USED TO ANALYZE CONCURRENT DOCUMENTATION

Variable 1 - Nature of Test - Substantive

Variable 2 - Nature of Test - Compliance

Variable 3 - Materiality - Was the account balance the audit procedure relates to material or immaterial?

Variable 4 - Reliance Placed - In recommending sample sizes for this audit procedure, are you placing some or no reliance on the system of internal accounting controls.

Variable 5 - Maximum Sample Size - The largest sample size you would recommend if all factors pointed towards a large sample.

Variable 6 - Minimum Sample Size - The smallest sample size you would propose if all factors pointed towards a small sample size and you still decide to perform the procedure.

Variable 7 - Experimental Treatment - Fair or strong as defined in Mock and Turner (1980).
Given the nature of the data which are mostly categorical, the general linear statistical modeling approach was deemed to be appropriate for assessing research questions 1, 2 and 3. Two special cases of the model were evaluated for relevance, ANOVA and regression, with the regression modeling approach being selected. The regression approach is a more general approach than ANOVA and can accommodate many independent variables, unequal sample sizes (for combinations of reference and demographic factors), and can be analyzed by computer programs primarily designed to evaluate the "importance of the variables" question implied by research question 2.

The basic model development approach consisted of three steps:

1. Fit the full model to the data assuming only main effects, that is, no interactions between the factors.

2. Based on the model in step 1, select those combinations of factors that provide the "best" reduced model, assuming only main effects, where "best" is defined as the model with the lowest Mallows Cp.\(^2\)

3. Fit the reduced model, assuming interactions. Using Mallows Cp, select those combinations of factors that provide the "best" reduced model.

These three steps were performed for each of the four audit procedures (E-5, E-6, E-9, E-10) over each of the two documentation approaches (retrospective, concurrent).

The hypothesized full model underlying the retrospective documentation task is:
\[ Y = \sum_{k=0}^{18} \beta_k x_{ik} + \epsilon_i \]

where \( x_{10} \leq 1, i=1, \ldots, n \), and where \( Y \) represents the observed sample size judgment, \( \beta_0 \) represents the combination of factors at their lowest level, \( \beta_1, \beta_2, \ldots, \beta_{18} \), represent cue weighting coefficients to be estimated for each factor, \( x_{11}, x_{12}, \ldots, x_{19} \), represent the internal accounting control reference factors in the experimental task, \( x_{19}, x_{110}, \ldots, x_{116} \), represent the demographic factors, and \( x_{118} \) represents the treatment effect factor. \( \epsilon_i \) is the statistical error term of the model. The correspondence of the model terms and reference and control factors is shown in Table 2.

The full hypothesized model underlying the concurrent documentation task (see Table 2) is:

\[ Y = \sum_{k=0}^{7} \beta_k x_{ik} + \epsilon_i \]

where \( x_{10} \leq 1, i=1, \ldots, n \) and where \( Y \) represents the observed sample size judgment, \( \beta_0 \) represents the combination of factors at their lowest level, \( \beta_1, \beta_2, \ldots, \beta_7 \) represent cue weighting coefficients to be estimated for each factor; and the \( x_{11}, x_{11}, x_{i2}, \ldots, x_{17} \), represent the factors referenced in the experimental task. \( \epsilon_i \) is the statistical error term of the model.

The hypothesized reduced models for both the concurrent and retrospective tasks will be of the general form:

\[ Y = \sum_{k=0}^{P-1} \beta_{ik} x_{ik} + \sum_{k=1}^{P-1} \beta_{ik+1} x_{ik+1} + \epsilon_i \]

where \( Y \) represents the observed sample size judgment, \( \beta_0 \) represents the combination of factors at their lowest level; \( \beta_1, \beta_2, \ldots, \beta_n \) represent cue...
Table 2

Relationship of Model Elements to Retrospective and Concurrent Documentation Factors

Retrospective Documentation

**Full Model**

\[
Y = \sum_{k=0}^{18} b_{ki} X_i + e_i
\]

- \(X_1\) = Audit Test Objectives Specified
- \(X_2\) = General and Specific Control References
- \(X_3\) = Cost Considerations
- \(X_4\) = Risk Considerations
- \(X_5\) = Other Audit Evidence
- \(X_6\) = Nature and Timing of the Test
- \(X_7\) = Nature of the Test-Substantive
- \(X_8\) = Nature of the Test-Compliance
- \(X_9\) = Compliance Test Results
- \(X_{10}\) = Years of Audit Experience
- \(X_{11}\) = Specialist Status
- \(X_{12}\) = Audit Level Training Courses Completed
- \(X_{13}\) = Firm Computer Audit Training
- \(X_{14}\) = Local Office Internal Control Training
- \(X_{15}\) = Local Office Statistical Training
- \(X_{16}\) = Commercial Experience
- \(X_{17}\) = Client Mix
- \(X_{18}\) = Experimental Treatment

Concurrent Documentation

**Full Model**

\[
Y = \sum_{k=0}^{7} b_{ki} X_i + e_i
\]

- \(X_1\) = Nature of the Test - Substantive
- \(X_2\) = Nature of the Test - Compliance
- \(X_3\) = Materiality of Account Being Audited
- \(X_4\) = Reliance Placed on System of Internal Accounting Controls
- \(X_5\) = Maximum Sample Size
- \(X_6\) = Minimum Sample Size
- \(X_7\) = Experimental Treatment
weighting coefficients to be estimated for each factor; \(X_{i1}, X_{i2}, \ldots, X_{im}\) represent the control reference, demographic, and treatment factors; \(X_{ik}X_{ik+1}\) represent the interaction factors and \(\varepsilon_i\) is the statistical error term of the model.

In conjunction with fitting the full and reduced model steps, all of the control reference factors, some demographic factors (retrospective task only), and the treatment factor are qualitative. This means that there is no statistical relation between these factors and the dependent variable \(Y\) in the regression models. Therefore, nothing can be stated regarding the type of statistical response function such as linear or polynomial. If only these qualitative factors are present in the reduced models, these regression models are analogous to ANOVA models. On the other hand, if some or all of the quantitative (interval scaled) demographic factors are represented in the model with the qualitative factors, these models become analogous to analysis of covariance (ANOCOVA) models provided certain assumptions are met. In the ANOCOVA models, there is a hypothesized statistical relation between the quantitatively scaled factors (covariates) and the dependent variable. A critical assumption of the ANOCOVA models is that of no interaction between the quantitatively scaled and qualitatively scaled factors. If these interactions are present, then separate regression models must be developed for each level of the quantitatively scaled factor. Thus, with reference to research question 1, there are a variety of potential models which may help understand auditor sample size judgments.

To further evaluate research question 1, steps 2 and 3 (see page 227) of the model building process must be undertaken. The goal of fitting
reduced models to the data is to find the best combination of control reference, demographic, and treatment factors that, in a statistical and judgmental sense, parsimoniously represent the data. Steps 2 and 3 were undertaken for each of the four audit procedures over each of the two tasks.

Research question 2, which is concerned with the relative importance of the final set of model factors, is answered by evaluating the reduced models developed in support of research question 1. Criteria for examining the relative importance of the factors include: marginal contribution of each factor to overall $R^2$ (percentage of variance explained); $t$-statistics for significance on each factor coefficient, and tolerance levels of each factor.\(^3\)

Research question 3 is concerned with the stability of the regression model factors and parameters over each of the four audit procedures (E-5, E-6, E-9, E-10) given a particular documentation approach (e.g., retrospective). Criteria utilized for evaluation of the results include: a comparison of the commonality of the control reference and demographic factors across procedures; an examination of the relative magnitudes of the regression model parameters, and an examination of the mathematical sign on the parameters, across the four audit procedures. If audit judgment models are the same across audit procedures, the same reference factors should be present in all procedures and the coefficients of the factors in the regression model should be of the same order of magnitude and display the same sign.

Results: Model Search and Development
Table 3 presents the summary of the results of reducing the "full" model considering only main effects (steps 1 and 2 of the previous section). As noted there, the variance explained in the models (adjusted $R^2$) is relatively low, but much higher than that found in Mock and Turner, even though in all cases the overall regression models are highly significant.

Table 4 shows the results of the analysis of the reduced models of Table 3, when interaction terms are built into the models. All procedures except E-9 retrospective and E-10 concurrent show significant increases (13% to 220%) in variance explained, and show a maintained or improved level of overall significance of the regression models.

Also, as may be noted in Table 3 and 4 and summarized in Table 5, the model elements do vary across audit procedures with few common model elements found across procedures. For example, for the retrospective documentation approach (see Table 5), "Nature and Timing" in both procedures E-5 and E-10 and its parameters are of the same order of magnitude and display the same sign. "Other audit evidence" (retrospective) appears in procedures E-5 and E-9 and its parameters are of the same order of magnitude and display the same sign across procedures. On the other hand, "Experimental Treatment" (retrospective) appears in procedures E-5, E-6, and E-10 and its parameters are of the same order of magnitude and display the same sign for procedures E-5 and E-10 but differ substantially for procedure E-6. In the instances where the same model elements are found over the procedures, the signs and relative magnitudes of the model parameters appear to be fairly stable, although not always, as discussed above.
Table 3
Summary Results of Main Effects, Full Model Reduction Based on Hallow's Cp Criterion for Four Audit Procedures Across Two Tasks

Retrospective Documentation Task

<table>
<thead>
<tr>
<th>Original Factors In the Model</th>
<th>Reduced Model Factors for Four Audit Procedures</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>E-5 (comparison) E-6 (pricing) E-9 (posting) E-10 (confirmation)</td>
</tr>
<tr>
<td>Control References</td>
<td>Regression Parameter</td>
</tr>
<tr>
<td>Audit Test Objectives</td>
<td>30.69</td>
</tr>
<tr>
<td>General and Specific Control References</td>
<td>34.83</td>
</tr>
<tr>
<td>Cost</td>
<td>-20.15</td>
</tr>
<tr>
<td>Risk</td>
<td>33.69</td>
</tr>
<tr>
<td>Other Audit Evidence</td>
<td>55.72</td>
</tr>
<tr>
<td>Nature and Timing of the Test</td>
<td></td>
</tr>
<tr>
<td>Nature of Test-Substantive</td>
<td></td>
</tr>
<tr>
<td>Nature of Test-Compliance</td>
<td></td>
</tr>
<tr>
<td>Compliance Test Results</td>
<td></td>
</tr>
<tr>
<td>Compliance Test Results</td>
<td>44.54</td>
</tr>
</tbody>
</table>
Table 3 (cont.)

Summary Results of Main Effects, Full Model
Reduction Based on Mallows' Cp Criterion
for Four Audit Procedures Across Two Tasks

Retrospective Documentation Task

<table>
<thead>
<tr>
<th>Original Factors in the Model</th>
<th>Reduced Model Factors for Four Audit Procedures</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>E-5 (comparison)</td>
</tr>
</tbody>
</table>

| Demographics | Regression Parameter | Standardized Parameter | Reg Parameter | STD Parameter | Reg Parameter | STD Parameter | Reg Parameter | STD Parameter | Reg Parameter | STD Parameter |
|--------------|----------------------|------------------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| Years of Audit Experience | 10.31 | .20 | -58.47 | .40 |
| Specialist Status | | | | |
| Audit Level Training | | | 39.28 | .27 |
| Firm Internal Control Training | | | |
| Local Office Internal Control Training | | | |
| Local Office Statistical Training | | | |
| Commercial Experience | -11.74 | -.18 | |
| Client Mix | | | |

| Treatment Effect | Experimental Treatment | -31.80 | -.27 | -31.56 | -.36 | -72.38 | -.23 |
|                  | $R^2$ | .15 | .40 | .13 | .43 |
|                  | Adjusted $R^2$ | .10 | .32 | .10 | .38 |
|                  | F-Statistic | 3.01 | 5.30 | 5.10 | 8.96 |
|                  | Significance | .02 | .00 | .01 | .00 |
Table 3 (cont.)

Concurrent Documentation Task

<table>
<thead>
<tr>
<th>Original Factors in the Model</th>
<th>Reduced Model Factors for Four Audit Procedures</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>E-5 (comparison)</td>
</tr>
</tbody>
</table>

Control Questionnaire Factors

<table>
<thead>
<tr>
<th>Nature of the Test-Substantive</th>
<th>Regression Parameters</th>
<th>Standardized Parameters</th>
<th>Reg Parameters</th>
<th>STD Parameters</th>
<th>Reg Parameters</th>
<th>STD Parameters</th>
<th>Reg Parameters</th>
<th>STD Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Materiality of Account</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reliance Placed on Controls</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum Sample Size Judgment</td>
<td></td>
<td></td>
<td>.19</td>
<td>.52</td>
<td>.35</td>
<td>.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minimum Sample Size Judgment</td>
<td></td>
<td>.52</td>
<td>.32</td>
<td>.49</td>
<td>.38</td>
<td>.51</td>
<td>.44</td>
<td></td>
</tr>
<tr>
<td>Nature of the Test-Compliance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Treatment Effect

<table>
<thead>
<tr>
<th>Experimental Treatment 12.28</th>
<th>.33</th>
<th>-15.25</th>
<th>-.27</th>
</tr>
</thead>
<tbody>
<tr>
<td>R²</td>
<td>.35</td>
<td>.59</td>
<td>.30</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>.28</td>
<td>.53</td>
<td>.27</td>
</tr>
<tr>
<td>F-Statistic</td>
<td>5.05</td>
<td>9.89</td>
<td>12.75</td>
</tr>
<tr>
<td>Significance Level</td>
<td>.01</td>
<td>.00</td>
<td>.00</td>
</tr>
</tbody>
</table>
Table 4

Summary of Reduced Model Parameters with Interaction Terms over Four Audit Procedures across Two Tasks

Retrospective Documentation

Procedure E-5

<table>
<thead>
<tr>
<th>Model Elements</th>
<th>Model Parameters</th>
<th>Standardized Model Parameters</th>
<th>2-tail Significance</th>
<th>Contribution to $R^2$</th>
<th>Tolerance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other Audit Evidence</td>
<td>-16.61</td>
<td>-.27</td>
<td>.01</td>
<td>.06</td>
<td>.82</td>
</tr>
<tr>
<td>Treatment Effect</td>
<td>-12.82</td>
<td>-.27</td>
<td>.06</td>
<td>.04</td>
<td>.78</td>
</tr>
<tr>
<td>Nature and Timing</td>
<td>21.31</td>
<td>.28</td>
<td>.03</td>
<td>.05</td>
<td>.62</td>
</tr>
<tr>
<td>Control Reference/Nature and Timing</td>
<td>47.22</td>
<td>.33</td>
<td>.02</td>
<td>.06</td>
<td>.50</td>
</tr>
<tr>
<td>Control Reference/Other Audit Evidence</td>
<td>47.64</td>
<td>.31</td>
<td>.01</td>
<td>.07</td>
<td>.72</td>
</tr>
<tr>
<td>Control Reference/Other Audit Evidence/Nature and Timing</td>
<td>-64.97</td>
<td>-.28</td>
<td>.06</td>
<td>.04</td>
<td>.47</td>
</tr>
<tr>
<td>Control Reference/Treatment Effect/Nature and Timing</td>
<td>-51.37</td>
<td>-.38</td>
<td>.00</td>
<td>.08</td>
<td>.51</td>
</tr>
<tr>
<td>Other Audit Evidence/Treatment Effect/Nature and Timing</td>
<td>77.65</td>
<td>.34</td>
<td>.01</td>
<td>.07</td>
<td>.59</td>
</tr>
<tr>
<td>Combination of Factors at their lowest Levels</td>
<td>78.59</td>
<td>1.306</td>
<td>.00</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

Overall Model E-5 Summary Statistics

Mean value of dependent variable - sample size judgment: 67.52
Coefficient of Variation of dependent variable: .89
Overall $R^2$: .39  F-Statistic: 4.53
Adjusted $R^2$: .31  Significance: .00
Improvement in Adjusted $R^2$ over main effects model: 220%
Table 4 (cont.)

Procedure E-6

<table>
<thead>
<tr>
<th>Model Elements</th>
<th>Model Parameters</th>
<th>Standardized Model Parameters</th>
<th>2-Tail Significance</th>
<th>Contribution to $R^2$</th>
<th>Tolerance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experimental Treatment</td>
<td>83.85</td>
<td>.96</td>
<td>.01</td>
<td>.05</td>
<td>.06</td>
</tr>
<tr>
<td>Nature of Test-Substantive</td>
<td>85.62</td>
<td>.90</td>
<td>.00</td>
<td>.09</td>
<td>.11</td>
</tr>
<tr>
<td>Nature of Test-Compliance</td>
<td>80.96</td>
<td>.89</td>
<td>.00</td>
<td>.12</td>
<td>.15</td>
</tr>
<tr>
<td>Audit Test Objectives/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nature of Test-Substantive</td>
<td>40.88</td>
<td>.24</td>
<td>.02</td>
<td>.04</td>
<td>.76</td>
</tr>
<tr>
<td>Audit Test Objectives/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audit Level Training</td>
<td>32.04</td>
<td>.37</td>
<td>.00</td>
<td>.09</td>
<td>.67</td>
</tr>
<tr>
<td>Other Audit Evidence/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nature of Test-Substantive</td>
<td>-42.18</td>
<td>-.24</td>
<td>.01</td>
<td>.05</td>
<td>.78</td>
</tr>
<tr>
<td>Nature of Test-Substantive/Nature and Timing</td>
<td>27.89</td>
<td>.20</td>
<td>.06</td>
<td>.03</td>
<td>.64</td>
</tr>
<tr>
<td>Nature of Test-Substantive/Experimental Treatment</td>
<td>-97.37</td>
<td>-.92</td>
<td>.00</td>
<td>.06</td>
<td>.07</td>
</tr>
<tr>
<td>Nature of Test-Compliance/Audit Level Training</td>
<td>-11.51</td>
<td>-.19</td>
<td>.09</td>
<td>.02</td>
<td>.62</td>
</tr>
<tr>
<td>Nature of the Test-Compliance/Experimental Treatment</td>
<td>-85.74</td>
<td>-.84</td>
<td>.01</td>
<td>.06</td>
<td>.08</td>
</tr>
<tr>
<td>Compliance Test Results/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental Treatment</td>
<td>-16.56</td>
<td>-.42</td>
<td>.02</td>
<td>.04</td>
<td>.22</td>
</tr>
</tbody>
</table>

Overall Model E-6 Summary Statistics

Mean value of dependent variable - sample size judgement: 63.60
Coefficient of variation for dependent variable: .69
Overall $R^2$: .57
Adjusted $R^2$: .49
F-Statistic: 7.33
Significance: .00
Improvement in Adjusted $R^2$ over main effects model: 53%
Table 4 (cont.)

Procedure E-9

<table>
<thead>
<tr>
<th>Model Elements</th>
<th>Model Parameters</th>
<th>Standard-ized Model Parameters</th>
<th>2-Tail Significance</th>
<th>Contribution to $R^2$</th>
<th>Tolerance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other Audit Evidence</td>
<td>-31.51</td>
<td>-.30</td>
<td>.01</td>
<td>.09</td>
<td>.99</td>
</tr>
<tr>
<td>Nature of Test-Substantive</td>
<td>34.73</td>
<td>.17</td>
<td>.14</td>
<td>.03</td>
<td>.99</td>
</tr>
<tr>
<td>Combination of Factors at Lowest Level</td>
<td>61.48</td>
<td>1.172</td>
<td>.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Overall Model E-9 Summary Statistics

Mean value of dependent variable -- sample size judgement: 44.86
Coefficient of variation in dependent variable: 1.17
Overall $R^2$: .13  F-Statistic: 5.10
Adjusted $R^2$: .10  Significance: .01
Improvement in Adjusted $R^2$ over main effects model: None
Table 4 (cont.)

Procedure E-10

<table>
<thead>
<tr>
<th>Model Elements</th>
<th>Model Parameters</th>
<th>Standardized Model Parameters</th>
<th>2-tail Significance</th>
<th>Contribution to $R^2$</th>
<th>Tolerance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nature and Timing</td>
<td>152.44</td>
<td>.30</td>
<td>.01</td>
<td>.05</td>
<td>.60</td>
</tr>
<tr>
<td>Experimental Treatment</td>
<td>-83.29</td>
<td>-.27</td>
<td>.01</td>
<td>.06</td>
<td>.80</td>
</tr>
<tr>
<td>Experience</td>
<td>-49.10</td>
<td>-.34</td>
<td>.00</td>
<td>.09</td>
<td>.78</td>
</tr>
<tr>
<td>Audit Level Training</td>
<td>48.65</td>
<td>.34</td>
<td>.00</td>
<td>.07</td>
<td>.64</td>
</tr>
<tr>
<td>Audit Test Objectives/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental Treatment</td>
<td>235.29</td>
<td>.40</td>
<td>.00</td>
<td>.03</td>
<td>.51</td>
</tr>
<tr>
<td>Audit Test Objectives/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nature and Timing/Audit Level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training</td>
<td>93.85</td>
<td>.42</td>
<td>.00</td>
<td>.08</td>
<td>.48</td>
</tr>
<tr>
<td>Experience/Audit Level Training</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>84.85</td>
<td>.67</td>
<td>.01</td>
<td>.05</td>
<td>.12</td>
</tr>
<tr>
<td>Experience/Audit Test Objectives/Audit Level Training</td>
<td>-31.34</td>
<td>.91</td>
<td>.00</td>
<td>.10</td>
<td>.12</td>
</tr>
<tr>
<td>Intercept</td>
<td>355.42</td>
<td>2.26</td>
<td>.00</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

**Overall Model E-10 Summary Statistics**

Mean value of dependent variable -- sample size judgement: 366.67
Coefficient of variation on dependent variable: .44
Overall $R^2$: .58  F-Statistic: 9.68
Adjusted $R^2$: .52  Significance: .00
Improvement in Adjusted $R^2$ over main effects model: 37%
Table 4 (cont.)

Concurrent Documentation

Procedure E-5

<table>
<thead>
<tr>
<th>Model Elements</th>
<th>Model Parameters</th>
<th>Standardized Model Parameters</th>
<th>2-tail Significance</th>
<th>Contribution to $R^2$</th>
<th>Tolerance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nature of Test-Substantive Experimental Treatment</td>
<td>-67.91</td>
<td>-.78</td>
<td>.02</td>
<td>.13</td>
<td>.21</td>
</tr>
<tr>
<td>Minimum Sample Size/Nature of the Test-Substantive</td>
<td>-25.82</td>
<td>-.69</td>
<td>.11</td>
<td>.04</td>
<td>.09</td>
</tr>
<tr>
<td>Minimum Sample Size/Experimental Treatment</td>
<td>1.14</td>
<td>.67</td>
<td>.01</td>
<td>.09</td>
<td>.21</td>
</tr>
<tr>
<td>Combination of Factors at Lowest Level</td>
<td>.47</td>
<td>1.21</td>
<td>.01</td>
<td>.14</td>
<td>.09</td>
</tr>
<tr>
<td>Overall Model E-5 Summary Statistics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean value of dependent variable -- sample size judgement:</td>
<td>70.91</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coefficient of variation on dependent variable:</td>
<td>.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall $R^2$:</td>
<td>.57</td>
<td>F-Statistic:</td>
<td>8.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted $R^2$:</td>
<td>.50</td>
<td>Significance:</td>
<td>.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Improvement in Adjusted $R^2$ over main effects model:</td>
<td>79%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4 (cont.)

Procedure E-6

<table>
<thead>
<tr>
<th>Model Elements</th>
<th>Model Parameters</th>
<th>Standardized Model Parameters</th>
<th>2-tail Significance</th>
<th>Contribution to R²</th>
<th>Tolerance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum Sample Size</td>
<td>.09</td>
<td>.24</td>
<td>.10</td>
<td>.04</td>
<td>.68</td>
</tr>
<tr>
<td>Minimum Sample Size</td>
<td>.60</td>
<td>.46</td>
<td>.00</td>
<td>.19</td>
<td>.88</td>
</tr>
<tr>
<td>Nature of the Test-Substantive/Maximum Sample Size</td>
<td>.25</td>
<td>.55</td>
<td>.00</td>
<td>.17</td>
<td>.58</td>
</tr>
<tr>
<td>Nature of the Test-Substantive/Minimum Sample Size/Maximum Sample Size/Experimental Treatment</td>
<td>-.005</td>
<td>-.29</td>
<td>.04</td>
<td>.06</td>
<td>.74</td>
</tr>
<tr>
<td>Combination of Factors at Lowest Level</td>
<td>17.29</td>
<td>.60</td>
<td>.01</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

Overall Model E-6 Summary Statistics

Mean value of dependent variable -- sample size judgement: 47.78
Coefficient of Variation on dependent variable: .60
Overall R²: .65  F-Statistic: 12.37
Adjusted R²: .59  Significance: .00
Improvement in Adjusted R² over main effects model: 11%
Table 4 (cont.)

Procedure E-9

<table>
<thead>
<tr>
<th>Model Elements</th>
<th>Model Parameters</th>
<th>Standardized Model Parameters</th>
<th>2-tail Significance</th>
<th>Contribution to $R^2$</th>
<th>Tolerance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum Sample Size</td>
<td>.35</td>
<td>.55</td>
<td>.00</td>
<td>.30</td>
<td>.10</td>
</tr>
<tr>
<td>Intercept</td>
<td>14.87</td>
<td>.28</td>
<td>.26</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

Overall Model E-9 Summary Statistics

Mean value of dependent variable -- sample size judgement: 51.71
Coefficient of variation on dependent variable: 1.01
Overall $R^2$: .30  
F-Statistic: 12.75
Adjusted $R^2$: .27  
Significance: .00
Improvement in Adjusted $R^2$ over main effects model: None
Table 4 (cont.)

Procedure E-10

<table>
<thead>
<tr>
<th>Model Elements</th>
<th>Model Parameters</th>
<th>Standardized Model Parameters</th>
<th>2-tail Significance to $R^2$</th>
<th>Contribution to $R^2$</th>
<th>Tolerance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum Sample Size</td>
<td>.51</td>
<td>.44</td>
<td>.01</td>
<td>.20</td>
<td>.99</td>
</tr>
<tr>
<td>Materiality</td>
<td>323.13</td>
<td>.34</td>
<td>.04</td>
<td>.11</td>
<td>.99</td>
</tr>
<tr>
<td>Intercept</td>
<td>-120.31</td>
<td>-.71</td>
<td>.45</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

Overall Model E-10 Summary Statistics

Mean value of dependent variable -- sample size judgement: 225.96
Coefficient of variation on dependent variable: .68
Overall $R^2$: .28 F-Statistic: 5.64
Adjusted $R^2$: .23 Significance: .01

Improvement in Adjusted $R^2$ over main effects model: None
Table 5

Summary of Main Effects Model Parameters
(Standardized) from Table 4 for
Retrospective and Concurrent Documentation

<table>
<thead>
<tr>
<th></th>
<th>Procedure-Retrospective</th>
<th>Procedure-Concurrent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>E-5</td>
<td>E-6</td>
</tr>
<tr>
<td>Nature and Timing</td>
<td>.28</td>
<td>--</td>
</tr>
<tr>
<td>Other Audit Evidence</td>
<td>-.27</td>
<td>--</td>
</tr>
<tr>
<td>Experimental Treatment</td>
<td>-.27</td>
<td>.96</td>
</tr>
<tr>
<td>Nature of Test-Substantive</td>
<td>--</td>
<td>.90</td>
</tr>
<tr>
<td>Nature of Test-Compliance</td>
<td>--</td>
<td>.89</td>
</tr>
<tr>
<td>Audit Level Training</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Experience</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Risk</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Maximum</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Minimum</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>
All of the resulting models over three procedures (E-5, E-6, E-9) for the retrospective documentation task appear to be regression analogues of the ANOVA model, that is, all factors in the model are qualitative. For procedure E-10, two interval scaled factors appear in the model and, thus, this model appears to be the regression analogue to the ANOCOVA model. Tests of the assumptions of the ANOCOVA model, particularly those of parallel response surfaces were negative indicating that significant interactions between qualitatively and quantitatively scaled factors were present. This means that separate regression response functions exist with different slopes and intercepts for each level of the quantitatively scaled factors.

The concurrent documentation task reflected one regression analogue of the ANOVA model, for procedure E-9. The other 3 procedures E-5, E-6, E-10 all appear to be regression analogues of the ANOCOVA model. Tests of the assumptions of the ANOCOVA models were negative indicating that these models all have different intercepts and slopes for each level of the quantitatively scaled factors: maximum and minimum sample size judgments. These results will be further discussed in the context of model interpretation which now follows.

Model Interpretability

The presence of some interaction terms in some of the models, although improving the variance explained and overall fit of the models, does cause some difficulty in explicitly interpreting the effects of the individual model elements. To aid in the interpretation of the various models, the following strategy is employed:
1. Use the reduced set of factors and analyze them through ANOVA techniques, which allow easier computation and comparison of group means for the procedures E-5, and E-6 for the retrospective documentation task.

2. Interpret model E-10 (retrospective), and models E-5, E-6, E-9, and E-10 of the concurrent task using the regression approach to model interpretation.

3. Interpret model E-9 (retrospective) using the ANOVA approach.

The results of these various interpretations are now discussed.

Procedure E-5--Packing Slip--Invoice Comparison--Retrospective Documentation. The model underlying procedure E-5 for the retrospective documentation method is comprised of two significant qualitative factors and five significant interaction terms. The model in notational form is:

\[ Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}X_{i2} + \beta_4 X_{i1}X_{i4} + \beta_5 X_{i1}X_{i1}X_{i4} + \beta_6 X_{i3}X_{i1}X_{i4} + \beta_7 X_{i1}X_{i2}X_{i4} + \epsilon_i \]

where \( Y_i \) = sample size judgment in the \( i \)th trial; \( \beta_0 \) = value of \( Y \) when qualitative factors are at the lowest level (zero); \( \beta_1, \beta_2, \ldots, \beta_7 \) are curve weighting coefficients; \( X_{i1} \) = other audit evidence, \( X_{i2} \) = experimental treatment, \( X_{i3} \) = control reference; and \( X_{i4} \) = nature and timing of the test, where \( i \) indicates the \( i \)th trial.

Since there are no quantitative factors, this model does not have a response function. The interpretation of the model is summarized in Table 6.
Table 6

Summary Interpretation of Model E-5—Retrospective Documentation

<table>
<thead>
<tr>
<th>Other Audit Evidence</th>
<th>Control Mental Reference</th>
<th>Nature Timing</th>
<th>Notational Model</th>
<th>Functional Model: Number Mean</th>
<th>Mode: Number sample size for judgment each sub-model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ref Ref Ref</td>
<td>Ref</td>
<td>E(Y) = β₀ + β₁ + β₂ + β₅ + β₆ + β₇</td>
<td>10.47</td>
<td>4</td>
<td>β₀ = 78.59</td>
</tr>
<tr>
<td>Ref Ref No Ref Ref</td>
<td>Ref</td>
<td>E(Y) = β₀ + β₁ + β₂ + β₄</td>
<td>76.80</td>
<td>8</td>
<td>β₁ = -16.61</td>
</tr>
<tr>
<td>Ref No Ref Ref Ref</td>
<td>No Ref</td>
<td>E(Y) = β₀ + β₁ + β₂ + β₇</td>
<td>104.87</td>
<td>4</td>
<td>β₂ = -12.82</td>
</tr>
<tr>
<td>Ref No Ref No Ref No Ref</td>
<td>Ref</td>
<td>E(Y) = β₀ + β₁ + β₂</td>
<td>48.16</td>
<td>1</td>
<td>β₃ = 47.22</td>
</tr>
<tr>
<td>No Ref Strong Ref Ref</td>
<td>No Ref</td>
<td>E(Y) = β₀ + β₂ + β₃ + β₆</td>
<td>61.62</td>
<td>1</td>
<td>β₄ = 28.64</td>
</tr>
<tr>
<td>No Ref Strong Ref No Ref</td>
<td>No Ref</td>
<td>E(Y) = β₀ + β₂</td>
<td>65.77</td>
<td>23</td>
<td>β₅ = -64.97</td>
</tr>
<tr>
<td>No Ref Strong No Ref Ref</td>
<td>No Ref</td>
<td>E(Y) = β₀ + β₂</td>
<td>65.77</td>
<td>5</td>
<td>β₆ = -51.37</td>
</tr>
<tr>
<td>No Ref Strong No Ref No Ref</td>
<td>No Ref</td>
<td>E(Y) = β₀ + β₂</td>
<td>65.77</td>
<td>26</td>
<td>β₇ = 77.65</td>
</tr>
</tbody>
</table>

As noted, several of the "sub-models" above are basically identical with respect to sample size judgments. Other sub-models exhibit wide variability, although the average or overall model does explain 32 percent of the variance. The majority of the auditors seemed to be affected by the experimental control treatment to the relative exclusion of the other factors (note in each case the effect (reduced sample size) as would be expected). There were, however, enough subjects exhibiting different underlying models to cause the overall model to consist of the combination of factors presented in Table 4.
Procedure E-6--Invoice Pricing Test--Retrospective Documentation.

The model underlying procedure E-6 for the retrospective task is the most complex (and difficult to explain) of all the models discussed in this paper. This complexity is due to the number of significant interaction terms in the model, some of which contain quantitatively-scaled factors interacting with qualitative factors. This model was restricted to main effects and two-way interactions only, due to the potentially large number of unexplainable higher-order interaction terms. Thus, the overall model may not "explain" as much variance as potentially possible. The model is, however, comparable to the other models for the other procedures in terms of explanatory power and significance, and thus is used in its present form. This model, in notational form, is represented as:

\[ Y_{i1} = \beta_0 + \beta_1 X_{i11} + \beta_2 X_{i12} + \beta_3 X_{i13} + \beta_4 X_{i14} + \beta_5 X_{i15} + \beta_6 X_{i16} + \beta_7 X_{i17} + \beta_8 X_{i18} + \beta_9 X_{i19} + \beta_{10} X_{i110} + \beta_{11} X_{i111} + \epsilon_i \]

where \( Y_{i1} \) = sample size judgment; \( \beta_0 \) = value of \( Y \) when the combination of factors are at their lowest level; \( \beta_1, \beta_2, \ldots, \beta_{11} \) = cue weighting coefficients; \( X_{i11} \) = experimental treatment, \( X_{i12} \) = nature of the test-substantive, \( X_{i13} \) = nature of the test-compliance, \( X_{i14} \) = audit test objectives, \( X_{i15} \) = audit level training, \( X_{i16} \) = other audit evidence, \( X_{i17} \) = nature and timing of the test, \( X_{i18} \) = compliance test results, \( \epsilon_i \) indicates the ith trial; and \( \epsilon_i \) is the statistical error term of the model.

The model elements and their relationship for the combination of the factors are not shown in this paper due to the complexity of the model and difficulty in presenting them. However, in referring to Table 4, it is
noted that the adjusted $R^2$ is .49 and the overall significance of the model is very good.

Procedure E-9--Posting Test--Retrospective Documentation. The model underlying procedure E-9 for the retrospective documentation is comprised of two qualitative factors as shown in Table 4. The resulting model, in notational form is:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i$$

where $Y_i$ = sample size judgment; $\beta_0$ = value of $Y$ when the qualitative factors are at their lowest level (zero); $\beta_1$, $\beta_2$ are cue weighting coefficients; $X_{i1}$ = qualitative factor—other audit evidence; $X_{i2}$ = qualitative factor—nature of the test—substantive, where $i$ indicates the $i$th trial.

Unlike the regression models underlying concurrent documentation procedures E-5, E-6, E-9, this regression model does not have a response function or significant interaction terms. Thus, the interpretation of the coefficients of the model is straightforward and is summarized in Table 7.

Table 7

Summary Interpretation of Model E-9—Retrospective Documentation

<table>
<thead>
<tr>
<th>Other Audit Evidence ($X_1$)</th>
<th>Nature of Test—Substantive ($X_2$)</th>
<th>Notational Model</th>
<th>Functional Model—Parameters</th>
<th>Model Mean Sample from Table 4 Judgment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference</td>
<td>Reference</td>
<td>$E(Y) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}$</td>
<td>64.70</td>
<td>$\beta_0 = 61.48$</td>
</tr>
<tr>
<td>Reference</td>
<td>No Reference</td>
<td>$E(Y) = \beta_0 + \beta_1$</td>
<td>29.97</td>
<td>$\beta_1 = -31.51$</td>
</tr>
<tr>
<td>No Reference</td>
<td>Reference</td>
<td>$E(Y) = \beta_0 + \beta_2$</td>
<td>96.21</td>
<td>$\beta_2 = 34.73$</td>
</tr>
<tr>
<td>No Reference</td>
<td>No Reference</td>
<td>$E(Y) = \beta_0$</td>
<td>61.48</td>
<td></td>
</tr>
</tbody>
</table>
As noted above and from Table 4, reference to other audit evidence results in a mean sample size reduction of 31.51, whereas reference to substantive nature of the test results in a mean sample size increase of 34.73.6

Procedure E-10--Accounts Receivable Confirmation--Retrospective Documentation. The group model underlying procedure E-10 for the retrospective documentation task has 2 significant qualitative factors, 2 significant quantitative factors, and 4 significant interaction terms.

The resulting model of Table 4 is shown in notational form as:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} X_{i2} + \beta_6 X_{i5} X_{i1} + \beta_7 X_{i3} X_{i4} + \beta_8 X_{i5} X_{i3} X_{i4} + \epsilon_i$$

where $Y_i$ = sample size judgment; $\beta_0$ = the intercept term; $\beta_1, \beta_2, \ldots, \beta_8$ are cue weighting coefficients, $X_{i1}$ = nature and timing of the test, $X_{i2}$ = experimental treatment effect, $X_{i3}$ = experience as an auditor, $X_{i4}$ = audit level training, $X_{i5}$ = audit test objectives, where $i$ indicates the $i$th trial.7

Since there are two quantitatively scaled factors which interact with each other, as well as the two qualitative factors, the change in mean sample size judgments is expressed in Table 8.8
Table 8

<table>
<thead>
<tr>
<th>xperience (X₁)</th>
<th>Audit Level Training (X₄)</th>
<th>Nature Test Objectives (X₅)</th>
<th>Audit and Test (X₂)</th>
<th>Treatment (X₃)</th>
<th>Change in Mean Response</th>
<th>Functional Change in Mean Response</th>
<th>Model Parameters from Table 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>unit inc</td>
<td>const ref</td>
<td>strong</td>
<td>β₃ + β₇X₅ + β₈X₄X₅</td>
<td>35.41 +</td>
<td>-34.35X₄</td>
<td>β₁ = 152.44</td>
<td>β₂ = -83.29</td>
</tr>
<tr>
<td>onst</td>
<td>unit increaser</td>
<td>strong</td>
<td>β₄ + β₇X₉X₅ + β₈X₄ + β₉X₃X₄</td>
<td>142.51 +</td>
<td>84.51X₅ +</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(-34.35)X₃X₄</td>
<td>-49.10</td>
<td>(-34.35)X₃X₄</td>
<td></td>
<td></td>
</tr>
<tr>
<td>unit inc</td>
<td>const ref</td>
<td>NR</td>
<td>β₃</td>
<td>48.66 +</td>
<td>84.51X₃ +</td>
<td>β₄ = 48.65</td>
<td>β₅ = 355.42</td>
</tr>
<tr>
<td>onst</td>
<td>unit increaser</td>
<td>strong</td>
<td>β₄ + β₇X₃ + β₈X₅X₄</td>
<td>35.41 +</td>
<td>(-34.35)X₃X₄</td>
<td>β₅ = 235.29</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(-34.35)X₃X₄</td>
<td>-49.10</td>
<td></td>
<td>β₅ = 84.85</td>
<td></td>
</tr>
<tr>
<td>constant</td>
<td>unit increaser</td>
<td>strong</td>
<td>β₄ + β₇X₃ + β₈X₅X₄</td>
<td>48.66 +</td>
<td>84.51X₃ +</td>
<td>β₅ = 93.85</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(-34.35)X₃X₄</td>
<td>-49.10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>onst</td>
<td>unit increaser</td>
<td>strong</td>
<td>β₄ + β₇X₃ + β₈X₅X₄</td>
<td>48.66 +</td>
<td>84.51X₃ +</td>
<td>β₅ = 31.34</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(-34.35)X₃X₄</td>
<td>-49.10</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviation key to above: inc = increase, const = constant, NR = no reference, ref = reference, incr = increase.

As can be seen, each combination of the factors in the model results in a variety of models with different intercepts and slopes. Here, as compared to some of the procedures of the concurrent documentation task, the magnitude of the changes in sample size judgments is quite great. As noted in Table 4, the overall model is highly significant and explains a significant amount (adjusted R² is .43) of the variance in auditor judgment.
Procedure E-5 Packing Slip-Invoice Comparison-Concurrent

Documentation Task. The model for procedure E-5 (concurrent documentation) as developed from Table 4, has two significant factors and two significant interaction terms. Since both interaction terms contain a quantitatively scaled factor, minimum sample size, the regression model is:

\[ Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{12} + \beta_3 x_{13} + \beta_4 x_{1i} x_{12} + \beta_5 x_{1i} x_{13} + \epsilon_i \]

where \( Y \) = sample size judgment; \( \beta_0 \) is the intercept term; \( \beta_1, \beta_2, \ldots, \beta_5 \) are cue weighting coefficients; \( x_{1i} \) = minimum sample size, \( x_{12} \) = nature of the task-substantive reference, \( x_{13} \) = experimental treatment effect, where \( i \) is the ith trial. This model has different response functions for the different "experimental treatment--nature of the test (substantive)" combinations. Also the differential effects of one qualitative factor on the intercept term depend on the particular class of the other qualitative factor.

The model elements are summarized in both notational and functional form in Table 9.
Table 9

Summary Interpretation of Model E-5--Concurrent Documentation

<table>
<thead>
<tr>
<th>Type of Experimental Treatment ($X_3$)</th>
<th>Nature of Test-Substantive ($X_2$) Response Function</th>
<th>Model Response</th>
<th>Functional Model Value</th>
<th>Model Value from Table 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strong</td>
<td>Substantive</td>
<td>$E(Y) = (\beta_0 + \beta_2 + \beta_3) + (\beta_4 + \beta_5)X_1$</td>
<td>$Y = -18.62 + 1.61X_1$</td>
<td>$\beta_3 = -25.8$</td>
</tr>
<tr>
<td>Fair</td>
<td>Substantive</td>
<td>$E(Y) = (\beta_0 + \beta_3) + \beta_5X_1$</td>
<td>$Y = 49.29 + 0.47X$</td>
<td>$\beta_2 = -67.91$</td>
</tr>
<tr>
<td>Strong</td>
<td>Compliance or Dual</td>
<td>$E(Y) = (\beta_0 + \beta_2) + \beta_4X_1$</td>
<td>$Y = 7.2 + 1.14X_1$</td>
<td>$\beta_4 = 1.14$</td>
</tr>
<tr>
<td>Fair</td>
<td>Compliance or Dual</td>
<td>$E(Y) = \beta_0$</td>
<td>$Y = 75.11$</td>
<td>$\beta_0 = 75.11$</td>
</tr>
</tbody>
</table>
The functional models imply that minimum sample size appears to serve as an anchor from which to modify sample size judgments based on combinations of the two qualitative factors internal control and nature of the audit procedure. Here the changes in sample size judgments across different combinations of the reference factors appear to be substantial.

Procedure E-6-Invoice Pricing Test--Concurrent Documentation. The derived model for procedure E-6 is a regression model with 2 significant factors and 2 significant interaction terms. The significant factors in this model are quantitatively-scaled and interact with the two qualitatively-scaled factors.

The model elements are represented in notation form as:

\[ Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i1} x_{i3} + \beta_6 x_{i1} x_{i2} x_{i3} x_{i4} + s_i \]

where \( Y_i \) = sample size judgment; \( \beta_0 \) = intercept term; \( \beta_1, \beta_2, \ldots, \beta_6 \) are cue weighting coefficients; \( x_{i1} \) = maximum sample size, \( x_{i2} \) = minimum sample size, \( x_{i3} \) = nature of the test-substantive, \( x_{i4} \) = experimental treatment effect, where \( i \) indicates the \( i \)th trial.

Since there are two important quantitatively-scaled factors in the model which interact with each other as well as the two qualitative variables, the change in mean sample size judgments is expressed in Table 10.
Table 10

Summary Interpretation of Model E-6—Concurrent Documentation

<table>
<thead>
<tr>
<th>Condition</th>
<th>Nature of Task—(X3)</th>
<th>Experimental Treatment (X4)</th>
<th>Change in Mean Response (Y)</th>
<th>Functional Change in Mean Response(Y)</th>
<th>Model Parameter</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>One unit increase Held at constant level</td>
<td>Substantive</td>
<td>Strong ( \beta_1 + \beta_5 X_3 + \beta_6 X_2 X_3 X_4 )</td>
<td>.34+((-0.005)X_2)</td>
<td>( \beta_1 = .09 )</td>
<td>( \beta_6 = .005 )</td>
<td>75.11</td>
</tr>
<tr>
<td>Held at Constant level</td>
<td>Substantive</td>
<td>Strong ( \beta_2 + \beta_5 X_3 X_4 )</td>
<td>.60+((-0.005)X_1)</td>
<td>( \beta_2 = .60 )</td>
<td>( \beta_5 = .25 )</td>
<td></td>
</tr>
<tr>
<td>One unit increase Held at constant level</td>
<td>Compliance or Dual</td>
<td>Strong ( \beta_1 )</td>
<td>0.09</td>
<td>( \beta_1 = .09 )</td>
<td>( \beta_5 = .25 )</td>
<td></td>
</tr>
<tr>
<td>One unit increase Held at constant level</td>
<td>Compliance &amp; Fair</td>
<td>Fair ( \beta_1 + \beta_5 X_3 )</td>
<td>.34</td>
<td>( \beta_1 = .09 )</td>
<td>( \beta_6 = .005 )</td>
<td>75.11</td>
</tr>
<tr>
<td>Held at constant level</td>
<td>Compliance or Dual</td>
<td>Strong ( \beta_2 )</td>
<td>.60</td>
<td>( \beta_2 = .60 )</td>
<td>( \beta_6 = .005 )</td>
<td></td>
</tr>
</tbody>
</table>

Mean sample size when all factors are zero: 75.11

As is noted in Table 10, each combination of the factors in the model results in a variety of models with different intercepts and slopes. In this case, the magnitude of the change in the sample size judgments is not very large. This implies that the mean sample size judgment of 75.11 (when all factor levels are zero) is relatively unaffected by the references (rationale) contained in the concurrent documentation. This, however, does not imply that the derived regression model is not sound. On the contrary, the derived model explains a significant amount (adjusted \( R^2 \) is .50) of the variation in auditor judgment, even though the magnitude of the variation is not great.
Procedure E-9-Posting Test--Concurrent Documentation. As developed from Table 4, one factor is present in the simple regression model underlying procedure E-9 (concurrent documentation). This is the ratio-scaled, maximum sample size factor. In notation form, this model is:

\[ Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i \]

where \( Y_i \) = sample size judgment; \( \beta_0 \) = the value of \( Y \) when \( X \) is zero; \( \beta_1 \) = the marginal contribution to \( Y \) of a unit increase in \( X \); \( X \) = maximum sample size judgment in the \( i \)th trial. The functional form of the model is:

\[ Y = 14.87 + .35X_{i1} \]

The interpretations of this simple regression model is straightforward. In effect, this model represents an anchoring and adjustment heuristic (maximum sample size). The marginal change in sample size judgment is .35 when the maximum sample size factor is varied by a unit amount. Although the marginal contribution to sample size is not great, the regression model is highly significant (\( d = .00 \)) and exhibits an adjusted \( R^2 \) of .27.

Procedure E-10--Accounts Receivable Confirmation Test--Concurrent Documentation. The model underlying procedure E-10 for the concurrent documentation task is comprised of one qualitative and one quantitative factor, as shown in Table 4. The resulting model, in notational form is:

\[ Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \]

where \( Y_i \) = sample size judgment; \( \beta_0 \) = the value of \( Y \) when the factors are at their lowest level; \( \beta_1, \beta_2 \) are cue weighting coefficients; \( X_{i1} \) = minimum sample size judgment; and \( X_{i2} \) = materiality factor, where \( i \) indicates the \( i \)th trial.
This model has a response function with a constant slope and intercept. The interpretation of the model parameters are summarized in Table 11.

Table 11

Summary Interpretation of Model E-10--Concurrent Documentation

<table>
<thead>
<tr>
<th>Minimum Sample Size ($X_1$)</th>
<th>Materiality Notational Model</th>
<th>Functional Model</th>
<th>Model Parameters from Table 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>One unit increase Material</td>
<td>$E(Y) = (\beta_0 + \beta_2) + \beta_1 X_1$</td>
<td>$202.82$</td>
<td>$\beta_0 = -120.31$</td>
</tr>
<tr>
<td>One unit increase Immaterial</td>
<td>$E(Y) = \beta_0 + \beta_1 X_1$</td>
<td>$-120.31$</td>
<td>$\beta_1 = .51$, $\beta_2 = 323.13$</td>
</tr>
</tbody>
</table>

As noted, since the slope and intercept of the response function are constant, the risk factor indicates how much the sample size judgment changes, given a constant level (anchor) for minimum sample size. The above model is highly significant ($\alpha = .01$), although not explaining a great deal of variance, with an adjusted $R^2$ of .23.

Model Testing - Predictive Ability

The ultimate assessment of the merit of a regression model is how well it fits new data. To assess the predictive ability of the models, cross validation techniques were employed.

The auditors' data were randomly assigned to two groups so that approximately equal numbers of auditors' data were in each group. One group of data was then used as a basis for developing the model and the second group of data was used as the cross validation set. A goodness-of-fit test based on the F-statistic was used to determine the significance or predictive validity of the model. The F-statistic used is
the ratio of the average squared residuals for auditors' data within the
cross validation set to the residual mean squares for the auditors' data
within the model development set.

The cross validation technique was performed for all audit
procedures (E-3, E-6, E-9, E-10) over both tasks. The summary results are
shown in Table 12. As noted there, all models over the two tasks have
high predictive ability significance. This implies that the models fitted
and presented in Table 4, are a reasonably good representation of the data
in that they have good ability to fit new data.

Model Representativeness - Nomothetic versus Idiographic

A major premise of this paper is the nomothetic view that a group
model of auditor judgment can be developed and defended. This nomothetic
view, in effect, assumes that significant individual differences, if at
one time existent, have been mitigated by training or other means. To see
how well this assumption holds, several analyses were made on a case by
case basis to determine the effects on the judgment models of extreme
departures by individual auditors from the group norms.

These analyses were directed to evaluating the sensitivity of the
regression judgment models to the data for each auditor. These techniques
included calculation of the deleted (Press) residuals, Mahalanobis
distance, and Cook's (1977) distance. The deleted (Press) residual is the
residual that would be obtained if an auditor's data were omitted from the
computation of the regression. Mahalanobis distance is the distance of
each auditor's data from the mean of all auditors' data used to estimate
the regression equation—a large distance indicates that the case is an
outlier in the space defined by the factors. Cook's distance is a measure
Table 12

Summary Statistics for Cross Validation of Models Underlying Four Audit Procedures over Two Tasks

<table>
<thead>
<tr>
<th></th>
<th>Retrospective</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>E-5</td>
<td>E-6</td>
<td>E-9</td>
<td>E-10</td>
</tr>
<tr>
<td>F-Statistic</td>
<td>5.96</td>
<td>16.60</td>
<td>2.26</td>
<td>2.61</td>
</tr>
<tr>
<td>Significance</td>
<td>.00</td>
<td>.00</td>
<td>.01</td>
<td>.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Concurrent</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>F-Statistic</td>
<td>3.01</td>
<td>2.98</td>
<td>2.10</td>
<td>4.87</td>
</tr>
<tr>
<td>Significance</td>
<td>.01</td>
<td>.01</td>
<td>.02</td>
<td>.00</td>
</tr>
</tbody>
</table>
of the change in the coefficient of the regression that would occur if
that auditor's data were omitted from the computation of the coefficients.

\[
\text{Cook's Distance} = \frac{\sum \left( \frac{r^2}{(p+1)\text{RMS}} \right)}{n}
\]

where \( r \) denotes the deleted (press) residual, where the term (+1) is
dropped if the intercept is zero, \( \text{RMS} \) denotes residual mean square and \( v \)
\( \omega \left( \frac{1}{N} + d \right) \) where the term \( \frac{1}{N} \) is dropped if there is a zero intercept and \( d \) is
the Mahalanobis distance divided by \( N-1 \) (or \( N \) if a zero intercept). \( \omega \) is a
weight assigned to each subject's data—assumed to be one unless
explicitly stated.

The results of these analyses are presented in Table 13. As noted
there, in most cases the relative magnitudes and signs of the coefficients
remain fairly consistent. There are, however, some exceptions where
outliers in the data cause serious changes in the coefficients, e.g., the
change of some of the interaction coefficients in E-10 (retrospective).

These outliers were not removed from the models presented in Table 4
since they represented valid representations of individual auditors in
the experimental procedures and were not errors in the data. The presence
of these outliers, where their presence causes serious changes in the
coefficients, gives some cause for questioning the complete satisfaction
of the nomothetic model assumption. Although not reported in this paper,
it was noted during the analysis that the subjects who exhibited
"non-normal" behavior in a given procedure such as E-10 (retrospective)
exhibited "normal" behavior in other audit procedures and conversely.

In general, the lack of homogeneity in aggregate or group model
building research, should not be troublesome as long as general group
tendencies of individual subjects are noted. In the current research.
Table 13
Comparison of Estimates of Regression Coefficients when Considering Cases with Largest Cook's Distance

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Model Element</th>
<th>Table 4 Model Parameter</th>
<th>Model Parameter Omitting Case with Largest Cook's Distance</th>
<th>Relative % Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>E-5</td>
<td>OAE</td>
<td>-60.59</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TRT</td>
<td>-30.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CR/</td>
<td>213.93</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NT</td>
<td>60.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CR/</td>
<td>60.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>OAE/</td>
<td>-174.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NT</td>
<td>-174.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CR/</td>
<td>-210.63</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TRT/</td>
<td>120.39</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NT</td>
<td>87.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-6</td>
<td>TRT</td>
<td>88.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NTS</td>
<td>85.62</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NTC</td>
<td>80.96</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ATO/</td>
<td>40.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NTS</td>
<td>32.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ATO/</td>
<td>-42.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ALT</td>
<td>27.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>OAE/</td>
<td>-97.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NTS</td>
<td>-11.51</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>NTS</td>
<td>-85.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TRT</td>
<td>-16.56</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 13 (cont.)

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Model Element</th>
<th>Table 4 Model Parameters</th>
<th>Model Parameter</th>
<th>Relative % Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Omitting Case</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>with Largest</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Cook's Distance</td>
<td></td>
</tr>
<tr>
<td>E-9</td>
<td>OAE</td>
<td>-31.51</td>
<td>-28.47</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>NTS</td>
<td>34.73</td>
<td>9.14</td>
<td>74</td>
</tr>
<tr>
<td></td>
<td>INT</td>
<td>61.48</td>
<td>59.60</td>
<td>3</td>
</tr>
<tr>
<td>E-10</td>
<td>INT</td>
<td>355.42</td>
<td>339.53</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>TRT</td>
<td>-83.29</td>
<td>-80.88</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>NT</td>
<td>152.44</td>
<td>156.27</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>ALT</td>
<td>48.66</td>
<td>51.09</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>ATO/NT/ALT/EXP/ATO/EXP/ATO/ALT/EXP/ATO/</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-31.35</td>
<td>634</td>
</tr>
<tr>
<td></td>
<td>TRT</td>
<td>235.29</td>
<td>66.90</td>
<td>72</td>
</tr>
<tr>
<td>E-5</td>
<td>TRT</td>
<td>-25.82</td>
<td>-30.53</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>NTS</td>
<td>-67.91</td>
<td>-63.44</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>MISS/NTS/MISS/TRNT</td>
<td></td>
<td>1.14</td>
<td>-4</td>
</tr>
<tr>
<td></td>
<td>INT</td>
<td>75.11</td>
<td>75.58</td>
<td>-2</td>
</tr>
<tr>
<td>E-6</td>
<td>MASS</td>
<td>.09</td>
<td>.01</td>
<td>89</td>
</tr>
<tr>
<td></td>
<td>MISS</td>
<td>.60</td>
<td>.69</td>
<td>-16</td>
</tr>
<tr>
<td></td>
<td>NTS/MASS/MISS/NTS/MISS/MASS/TRNT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>INT</td>
<td>17.29</td>
<td>17.43</td>
<td>-1</td>
</tr>
<tr>
<td>E-9</td>
<td>MASS</td>
<td>.35</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>INT</td>
<td>14.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-10</td>
<td>MISS</td>
<td>.51</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>RISK</td>
<td>323.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>INT</td>
<td>-120.31</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
KEY to Abbreviations in above table elements.

AE = Other Audit Evidence
TRI = Experimental Treatment
GR = General and Specific Control Reference
NT = Nature and Timing of the Test
INT = Intercept of the Model
NTS = Nature of the Test - Substantive
NTC = Nature of the Test - Compliance
ATO = Audit Test Objectives
ALT = Audit Level Training Courses Completed
CTR = Compliance Test Results
EXP = Years of Audit Experience
MISS = Minimum Sample Size
MASS = Maximum Sample Size
with minor exceptions, the nomothetic model assumption seems to be reasonable and adequately reflects the data. Another evidence of this conclusion is the high predictive ability of all the models even when some of the coefficients of some models may vary in the presence of "non-normal" auditor behavior.

Documentation Effects

To this point in the paper, the four audit procedures have been compared within the two documentation tasks. Research question six is concerned with a comparison of the two groups of task models, that is, the retrospective documentation models versus the concurrent documentation models.

As described in the introduction of this paper, the concurrent documentation task consisted of structured documentation judgments, whereas the retrospective documentation task consisted of open-ended, rationale memos developed after the judgments were made. As also previously stated, the basis difference between these two documentation tasks was one of timing. The question of interest now is directed to an assessment of which of the two types of documentation provides the better models in context of overall variance explained in auditor judgment.

Since the model elements are somewhat different for the two types of models, the typical methods of comparing regression models cannot be utilized. Rather, a more subjective appraisal is made based on looking at general model statistics for the two types of models. These statistics include: Adjusted $R^2$, Residual Mean Squares, and Standard Error of the Estimate. The results of these statistics are summarized in Table 14.
Table 14

Summary of Statistics to Compare Retrospective (Retro) Versus Concurrent (Conc) Documentation Models over Four Procedures

<table>
<thead>
<tr>
<th>Statistics</th>
<th>E-5 Retro</th>
<th>E-6 Retro</th>
<th>E-6 Conc</th>
<th>E-9 Retro</th>
<th>E-9 Conc</th>
<th>E-10 Retro</th>
<th>E-10 Conc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adjusted $R^2$</td>
<td>.32</td>
<td>.50</td>
<td>.49</td>
<td>.59</td>
<td>.10</td>
<td>.27</td>
<td>.52</td>
</tr>
<tr>
<td>Residual Mean Square</td>
<td>2458</td>
<td>838</td>
<td>985</td>
<td>337</td>
<td>2472</td>
<td>1996</td>
<td>11911</td>
</tr>
<tr>
<td>Standard Error of the Estimate</td>
<td>50</td>
<td>29</td>
<td>31</td>
<td>18</td>
<td>50</td>
<td>45</td>
<td>109</td>
</tr>
</tbody>
</table>

Best Model in Terms of Above 3 Criteria (X):

<table>
<thead>
<tr>
<th></th>
<th>E-5</th>
<th>E-6</th>
<th>E-9</th>
<th>E-10</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>
As noted in Table 14, the concurrent documentation models appear to be superior to the retrospective documentation models for all audit procedures except for E-10. The models for procedure E-10 support the retrospective task as being superior. One explanation for this is that the three procedures, E-5, E-6, E-9, are by their nature more structured, and perhaps require more specific reliance on or reference to specific kinds of documentation. On the other hand, procedure E-10 is concerned with confirmation of accounts receivable and thus, by nature this procedure may be more general and thus capture the more general kinds of control references demonstrated in the models of Table 4.

Discussion of Results

The analytical results presented in the previous sections are addressed to two general research issues. Research questions 1 through 5 address the significant amount of auditor judgment variability which was not "explained" by the experimental treatments in the Mock and Turner study. Research question 6 is concerned with the comparative "value" of two methods of documenting auditor judgment.

Unexplained Variability

Using statistical methods based on the general linear model, the first five research questions were addressed in an attempt to derive improved descriptive models of auditor judgment. In comparison to the variance explained by experimental treatments alone, models based upon data developed from both concurrent and retrospective documentation were derived which exhibited greater explained variance (adjusted $R^2$'s). In addition to explaining more variance, these models were found to be robust both with respect to predictive validity, and various residual analyses.
Although some evidence of individual specific (idiosyncratic) behavior was obtained, general (homothetic) models were found to describe a substantial amount of the observed variance in auditor judgments.

A previous section presented the various models in detail including a discussion of the functional form of the models, the contribution of each factor, and the signs and weights of included variables. In general, these models were found to be quite complex with several models containing significant interaction terms. In most cases the included variables and their direction of impact were as expected. For example, reliance on other audit evidence was negatively related to sample size recommendations. In only one case did the addition of training or other demographic variables result in a significant increase in explained variance (as a main effect). On the other hand, several variables which may have been expected to appear in the models (e.g., cost and risk considerations) did not appear.

The derived judgment models for all audit procedures given concurrent documentation, may be interpreted as anchoring and adjustment heuristics based on either a minimum or maximum sample size judgment. 11

All results taken into consideration, the combined approach of the multiple research methodologies utilized in the Mock and Turner study, and the multivariate statistical methods applied in this paper have led to improved descriptive models of audit sample size judgments.

Comparative "Value" of the Documentation Methods

Previous analyses in Mock and Turner had shown that when documentation method was treated as an experimental variable in an ANOVA
model, no significant differences in the recommendations analyzed here were observed. This left open the question of whether one method of documenting auditor judgments was better than another form either a research or audit review perspective. Although both the open-ended (retrospective) and structured (concurrent) methods were found lacking in comprehensiveness in Mock and Turner, careful study and coding of the documentation produced by both methods led to models with statistically high descriptive and predictive characteristics. In the cases of the transaction cycle audit procedures (E-5, E-6, and E-9), the structured, concurrent method produced a greater amount of explained variance. In the case of a test of account balances, the open-ended, retrospective method led to higher explained variance. Whether the two approaches differ in terms of cost of coding and analysis or produce models which are generally descriptive of auditor behavior is the subject of future research.
1. Mock and Turner (1980) used two experimental treatments. Only the Strong-Fair internal control treatment is analyzed as a treatment effect in this study. The second treatment (guidance/approach) is evaluated in terms of auditor rationale documentation in this paper.

2. The regression modeling approach was used to determine the nature of the auditor decision models by evaluating the "best set" of factors to be included. Best was defined in terms of Mallows' CP (Daniel and Wood, 1971, p. 86):

\[
CP = \frac{RSS}{s^2} - (N-2p')
\]

where RSS is the residual sums of squares for the best subset being tested, \( p' \) is the number of factors in the subset (including the intercept, if any), \( s^2 \) is the residual mean square based on the regression using all independent variables. Best is defined as the smallest CP. The t-statistics for the coefficients of the factors for the subset that minimizes \( CP \) tend to be greater than 2 in absolute value. In the language of stepwise regression, the subset that minimized \( CP \) is such that the F-to-remove value for variables in the subset tend to be greater than two and the F-to-enter values for the remaining variables tend to be less than two.
3. In studies such as this where there are a large number of non-orthogonal independent variables (factors) multicollinearity among the independent variables is common. This creates a problem in that the model elements and predictive ability of the model are unstable. Tolerance is a measure of the degree of multicollinearity of a given factor with all other factors (independent variables) in the model. Tolerance values approaching 1 indicate no multicollinearity and values approaching zero indicate high multicollinearity.

4. In an attempt to make important interactions unimportant, several transformation of the dependent variable (sample size judgments) were undertaken. These included log, square root and squared transformations. Although removing some important interactions and making them unimportant, the transformations allowed additional interaction terms to enter the model. In the case of the squared transformation, all models showed significant improvements in $R^2$ but at the sacrifice of interpretability. Hence, these transformed results are not reported.

5. Table 3 shows eight main effects for this model. A full factorial design would have potentially resulted in 256 model elements, most of which would have been uninterpretable.

6. These sample size increases are reported here as point estimates for completeness these should have confidence intervals placed around them using the Bonferroni interval.
7. Note that this is the only model that contains demographics as main effects.

8. The change in mean response is used here and for Procedure E-6 (concurrent) for simplicity in representing the more complex, interactions of the various submodels. The change function is a partial derivative of the overall model based on the quantitative factor held constant in the model.

9. This factor was not significant in the overall model independently.

10. These terms were not significant in the model but do appear in interaction terms and hence are included for completeness.

11. Given that some audit firms have established minimum sample size standards for certain audit circumstances, such standards can influence the risk of unwarranted reliance and of undetected errors and irregularities.
REFERENCES


