AUDITING RESEARCH SYMPOSIUM

Editors: A. Rashad Abdel-khalik and Ira Solomon

We gratefully acknowledge the financial support received for this symposium:

The Price Waterhouse Foundation sponsored the symposium.

The Arthur Andersen & Co. Alumni Professorship underwrote the printing and dissemination of the proceedings.

Office of Accounting Research, University of Illinois at Urbana/ Champaign
Contents

Preface / vii

Address
Technical Proficiency Is Just the Beginning / 1
Joseph E. Connor

Session One
Model-based Sampling and Its Implications for Auditors / 15

Discussants
David Burgstahler / 47
James K. Loebbecke / 54

Reply / 58

Session Two
On the Fundamental Nature of Professional Opinions: The Traditional, Bayesian, and Epistemc Methods of Attestation / 63
Raymond Dacey and Bart Ward

Discussant
William R. Scott / 92

Session Three
Models for Multilocation Audits / 99
Steve J. Aldersley and Donald A. Leslie

Discussants
Paul J. Beck / 129
James L. Kiriland / 134
Session Four

Estimation of Errors in Financial Statements / 139
  John J. Willingham and William F. Wright

Discussants
  Carl S. Warren / 168
  Philip D. Wedemeyer / 174

Session Five

Evidence on the Nature of Professional Judgment in Public Accounting / 181
  Michael Gibbins and Craig Emby

Discussants
  Carol E. Eger / 213
  James C. Cook / 222

Session Six

The Demand for Quality-Differentiated Audit Services in an Agency-Cost Setting: An Empirical Investigation / 229
  Zoe-Vonna Palmrose

Discussants
  Nicholas Depuch / 253
  Robert K. Elliott / 264

Reply / 271

Session Seven

  William R. Kinney, Jr.
Participants

Abraham D. Akresh
Steve J. Aldersley
Andrew D. Bailey, Jr.
Michael J. Barrett
David Burgstahler
Jane E. Campbell
Joseph E. Connor
James C. Cook
Raymond Dacey
Nicholas Doupch
Gordon L. Duke
Carol E. Eger
Robert K. Elliott
Craig Emby
Barry Epstein
William L. Felix, Jr.
Marguerite H. Fisher
Michael Gibbins
Richard A. Grimlund
Sabry Heakel
James A. Heintz
Gary L. Holstrom
Janet T. Kimbrell
William R. Kinney, Jr.
James L. Kirland
Chen-en Ko
Jack L. Krogstad
W. Morley Lemon
Barry L. Lewis
Donald A. Leslie
James K. Loebbecke
Robert G. May
William F. Messier, Jr.
Jane E. Mutchler
Christopher J. Nachtsheim
Zoe-Yonna Palnrose
Kurt Pany
David Pearson
Larry E. Rittenberg
Donald M. Roberts
Joseph J. Schultz, Jr.
Mary S. Schroeder
William R. Scott
David P. Shields
Laventhal & Horwath
Clarkson Gordon
University of Minnesota
University of Illinois, Chicago
University of Washington
Ohio State University
Price, Waterhouse & Co.
Ernst & Whinney
University of Idaho
Washington University
University of Minnesota
Stanford University
Per, Maruick, Mitchell & Co.
University of Alberta
Blackman, Kallick, & Co.
University of Arizona
Purdue University
University of Alberta
University of Iowa
McGladrey, Hendrickson & Pullen
Indiana University
University of Southern California
Oklahoma State University
University of Michigan
Deloitte Haskins & Sells
University of Minnesota
Creghtan University
University of Waterloo
University of Pittsburgh
Clarkson Gordon
University of Utah
University of Texas
University of Michigan
Ohio State University
University of Minnesota
University of California
Arizona State University
Arthur Young & Co.
University of Wisconsin
Arthur Andersen & Co.
Arizona State University
University of Houston
University of Waterloo
Rice University
University of Illinois Accountancy Faculty

A. Rashad Abdel-khalik
Paul J. Beck
Norton M. Bedford
Maureen H. Berry
Richard Boland
Gerald D. Brighton
Clifton E. Brown
John S. Chandler
Nancy A. Desmond
Philip E. Fess
Thomas J. Frecka
Ralph Greenberg
H. Peter Holzer
Karen Hreha
Robert W. Jamison
Orace E. Johnson
Rafael Lazimy

James C. McKeown
H. Kent McMath
Frederick L. Neumann
Denise Nitterhouse
Soong H. Park
Kenneth W. Perry
H. M. Schoenfeld
Florence Sharp
Robert F. Sharp
Peter A. Silhan
Charles H. Smith
Ira Solomon
Rick Wilberg
Eugene Willis
Richard E. Ziegler
Vernon K. Zimmerman

University of British Columbia
University of Texas
University of Oklahoma
University of Georgia
Arthur Andersen & Co.
Peat, Marwick, Mitchell & Co.
Peat, Marwick, Mitchell & Co.
Louisiana State University
Boston University
University of Minnesota
This Auditing Research Symposium series began ten years ago in the Department of Accountancy of the University of Illinois. The present volume contains the papers and discussant comments which comprised the sixth of the series held November 9–10, 1984.

As has been the custom, we did not select papers that espouse any particular philosophy or approach. Rather the symposium was open to all papers as long as the conclusions contained therein were supported by the underlying research. The quality of the manuscripts, as in previous years, was the discriminating variable in the selection process.

New this time, however, was that each submitted manuscript was anonymously refereed. Further, authors of accepted papers were asked to make revisions suggested by the referees. We believe that this process was effective in generating papers of higher quality than that typically found in published conference research reports. The willingness of the authors to submit their papers to the referee process and the diligent efforts of the referees were necessary for it to be successful. Consequently, we extend our thanks and appreciation for work well done.

The symposium also benefited from the time and efforts of the discussants who often introduced alternatives to the ideas that were presented in the papers and, in general, facilitated the creation of an environment within which there could be healthy intellectual discourse. Invited participants also contributed to this environment. Thus, to both the discussants and invited participants, we express our sincere gratitude.

The extensive work this symposium required could not have been accomplished without the support of the troops working behind the scene. We extend our appreciation, therefore, to Phyllis Sloat and Alice Waldoff.

Finally, we trust that the works contained in this volume will be of interest to auditing researchers and practitioners. We are pleased to have been a part of it.

A. Rashad Abdel-khalik
Ira Solomon
Co-Editors
Most of you are educators. The accounting graduates of the institutions you represent are the finished products of your efforts and the raw materials of the practicing profession. As a practitioner, I have a vested interest in your activities because the fate of the profession rides almost solely on the quality of its people.

Like other human activities, professional accounting stands perpetually on the edge of change, and you, as educators, have the unenviable task of trying to discern its direction. Because today's entrants will spend their professional careers in an uncertain future, how to prepare them is a major conundrum.

My comments this evening are one practitioner's view of what the signals coming from today's environment seem to be saying about professional qualifications. These, in turn, may shed some light on the kind of accounting and auditing education needed.

What are the primary characteristics of the environment in which today's professionals practice?

Environmental Condition No. 1 — higher expectations. Professional accounting has come out of the closet; the press has discovered that professional activities and accounting issues make news. Heightened
public awareness has been followed by higher expectations of the audit role from a public quick to register its outrage — and to bring court action — whenever it appears that those expectations have not been met.

Environmental Condition No. 2 — competition. The fierce competition that is an indisputable fact of professional life has two faces — good and bad. The good face is seen in services more responsive to user needs. The bad face appears in charges that the profession has sold out to the gods of commerce. The false assertion that audit services are a commodity to be acquired based solely on price may be fueling competitive fires. Far worse, the competitive environment appears to be encouraging marginal auditing and promoting an increase in “shopping” for accounting principles.

Environmental Condition No. 3 — rule-based standards. The efforts of standards setters and regulators to raise the level of financial reporting have kept their printing presses running overtime. With a standards-setting philosophy running strongly in the direction of detailed recipes for every situation and away from standards that require judgment to apply, some suggest that professional requirements have been lowered. How much technical expertise is needed to read a rulebook?

Environmental Condition No. 4 — audit methodology. The advent of computers has changed — and will continue to change — audit methodology. In the process, the traditional staff-to-partner pyramid is narrowing at the base.

Environmental Condition No. 5 — audit versus nonaudit services, relates importantly to career opportunities available in our profession. As nonaudit services climb as a proportion of total professional services, audit services suffer an equal decline in percentage points. Analysts of the profession cry that auditing is no longer a growth industry.

Let’s look at these conditions — some of which are interrelated — in the context of people needs.

HIGHER EXPECTATIONS

In the past two years the press has had a field day probing the financial reporting implications of a number of spectacular business failures, with some of the more serious involving financial institutions. Adding more targets for reporter’s arrows, investigations by the SEC’s Enforcement Division have revealed far too many instances of alleged financial reporting shenanigans ranging from deliberate fraud
to misapplied accounting principles. For the second time in less than ten years, the profession is about to be subjected to a full scale Congressional investigation.

Looking at the evidence presented by the shabby reporting that allegedly occurred in many collapsed businesses and by the SEC’s charges of “cooked” books and “cute” accounting, it is possible to point an accusatory finger at nearly every participant in the financial reporting process — at preparers, at auditors, at the FASB, at the AICPA, and at the SEC through its oversight role.

Among these, the audit role is the most tenuous. Like it or not, auditors have a serious image problem. The public believes that the auditor’s sole mission in life is to prevent the publication of misleading financial statements — from whatever source, outright fraud or upside-down accounting. In far too many cases in the roster of evidence, financial statements bearing the audit stamp of approval were later, sometimes within a matter of weeks, revealed to be seriously defective.

Why? The discrete reasons vary case by case. Included among them are instances where the auditor complied with the exact letter of existing auditing standards but extenuating circumstances existed. For example, carefully considered judgments failed to withstand SEC scrutiny, or management collusion successfully buried the real facts.

But the public isn’t interested in alibis or in explanations that “the auditor CAN’T” protect the public from some class of events. The public expects performance — performance consistent with the public responsibilities auditors have elected to undertake.

True, the aggregate of proven and alleged audit failures is imperceptible when compared to the thousands of unassailable audits conducted annually. Yet, this impressive record is never reported; public judgments necessarily are based on the sordid events making headlines.

Is increased federal regulation the answer? More and tighter standards? Stiffer sanctions? Definitely not. The net of federally imposed rules and sanctions already in place did not stop the observed occurrences. The answer must come from within a profession that fully recognizes the scope of its public responsibilities and practices its professionalism on every engagement, large or small. Quite simply, we must improve our performance. We must shoot for zero failure.

And how does this environmental condition affect you role? Public expectations — and the knee-jerk reaction to seek redress for injuries attributed to alleged audit failure that flows from them — have clear implications for staff requirements. This condition raises those requirements and increases the burden on you to find ways to improve
the competency level of those emerging from your programs. We cannot afford to gamble with the ill-trained, the less than dedicated, with those who don't understand or appreciate the auditor's public mission and the responsibilities that go with it. The stakes are too high.

COMPETITION

I view competition as a healthy development — a challenge to be met. It wakes us up, strips away any aura of complacency that inhibits forward progress. Competition forces us to examine critically the market for our services and to ensure that services offered are fully in tune with market needs. Responding to real-world needs is what professional services are all about.

The roots of competition among professional firms are primarily economic. In common with other goods and services, the price of auditing services, driven by inflation, has risen dramatically over the past several years. Services buyers, also pinched by spiraling costs, have been reviewing all purchases, including audit services, to ensure that they are cost-effective.

That is understandable. I certainly do not want to be on the receiving end of goods or services that are overpriced. And I do not want to be accused of providing services that are not considered worth every penny the firm collects.

Nevertheless, the evils, real or perceived, attributed to competition's bad face must be recognized and addressed.

A totally false concept has emerged from the competitive fray and may be encouraging it. The notion that professional services are a commodity — fungible goods to be acquired based solely on price — is heard with irritating frequency.

The commodity change appears to spring from observable similarities among the major firms — similarities of size, of geographic coverage, of range of services offered, and of claims to superior levels of audit efficiency and effectiveness. All firms must comply with common accounting and auditing standards and with SEC regulations. Members of the SEC Practice Section of the AICPA Division for CPA firms must undergo triennial peer reviews and comply with numerous other requirements designed to promote high-quality practices.

But the sameness stops there. The existence of standards and of similar resources does not imply that the delivery of professional services has been reduced to a rote exercise. These are signs of a maturing profession, signs that numerous forces have acted to raise
the overall level of professional practice. Those of us in public practice know where we stand — and where others stand — in the quality pecking order that exists within the profession.

Charges that competition is encouraging marginal auditing and "principle shopping" are serious danger signals for the profession.

In the area of marginal auditing, consider this statement attributed by Time magazine to John Fedders, the chief of the SEC’s Enforcement Division: "There’s tremendous pressure on accounting firms to lower their prices. Some firms can, but only by cutting corners."

The impression, coming from a government official in possession of hard evidence, that some auditors are cutting corners is indeed a severe indictment of professional performance. At the same time, it yanks the underpinnings from the assertion that professional services are all the same.

"Principle shopping" — finding an auditor willing to support a non-GAAP accounting position that will improve the "shopper’s" reported earnings — is a product not only of competition but of the complex rule-based environment in which we practice. A successful "shopping" expedition pairs abdication of preparer responsibility with auditor betrayal of a public trust in a conspiracy to subvert the financial reporting system. An auditor approving suspect accounting to gain a competitive advantage is an auditor stripped of any trace of professionalism.

SEC action could end this insidious practice by denying the culprit any hope of financial gain. Labeling clear cases of "shopping" as such in the 8-K letter required of discharged accountants would serve as a red flag to a Commission already deeply concerned about the incidence of "shopping." In my view, whenever the SEC finds that the audit stamp of approval has been placed on the auction block, the Commission should refuse to accept continuance of the supplanting accountants, thus ensuring that unacceptable conduct does not breed financial rewards.

There is no truth to the assertion that competition necessarily forces out professionalism. Those who sacrifice their professionalism for commercial advantage undoubtedly would do so regardless of the prevailing competitive climate. Competition need not alter our priorities: providing the highest quality of professional services, maintaining independence and objectivity, and fulfilling our responsibilities to the public. Our professionalism is our service. Anything less is worthless — at any price.

Active competition affects the educator's role in much the same way as heightened public expectations. Active competition, in and of
itself, places a premium on staff quality. We must anticipate that the services buyer interested in getting full mileage from the company's audit dollars will be watching for any evidence of imperfect performance. We must be sure that all work performed has a legitimate purpose, and that it is done effectively — the first time. All those assigned to the engagement — from the lowest to the highest — must be knowledgeable, alert professionals fully attuned to the client's service needs. Any evidence to the contrary will swiftly and surely make its way into the potent underground information conduit where it can be expected to work to an audit firm's disadvantage in any number of unrelated engagements.

METHODS OF COMPETITION

The ways professional firms prepare to meet competition shed further light on our personnel requirements and on your role in helping us to fill them. Let's consider three activities that help firms to compete: selecting an organizational structure, expanding geographic coverage, and improving technical quality.

In these activities — indeed, in all aspects of professional accounting — the performance of individuals is of paramount concern. A firm's national and international credentials carry weight in the auditor-selection process. But once a choice is made, a handful of individuals carry the load in getting the job done in accordance with promised specifications. It is individuals who deliver services, whose actions preserve and improve upon a firm's reputation or damage it.

A firm's organizational structures makes a difference. Price Waterhouse, for example, has a decentralized structure. We are strongly committed to the principle that the ultimate responsibility for delivery of quality services to each client rests with the firm owner — the partner — who deals with that client. The alternatives are varying degrees of centralization.

The selection of a structure that permits binding decisions to be made at the client's conference table versus one that requires referral through a multilevel chain of command affects the quality and character of service and the quality of people needed to supply it. The risks endemic to our profession demand that decision-making authority be entrusted only to individuals with a proven depth of technical expertise, experience and business judgment necessary to make reasoned decisions — and the good sense to consult where consultation is needed.

Professional firms evaluate potential new markets before commit-
ting resources to gain entrance. New offices may be opened to provide better service to an existing client base or to build a practice from scratch in a promising area. Once the shingle is hung, the real work begins. Building contacts and a reputation for dependability and quality, letting business know the range of services available, is a long, slow process. Again, success depends on individuals — individuals who are resourceful business people with vision and entrepreneurial spirit.

To be responsive to client service needs, professional firms must stay on the cutting edge of technical change. Monitoring the accounting standards-setting process and accounting developments originating in Washington — and providing input where desirable — is essential to the quality of basic accounting services.

Commonly, these activities are the responsibilities of a few knowledgeable individuals. They take the lead in marshaling the firm’s resources to develop reasoned positions based on experience with accounting’s effect on business. Not content to watch the parade from the sidelines, they participate actively in shaping the direction of future policies and standards.

The qualifications for such activities underscore the importance of your role in equipping entry-level professionals with the broad-based foundation they will need in today’s practice environment.

Professionals first and foremost, the individuals competing in these areas must be skilled in the art of persuasion, skilled in written and oral communications. They must be first-rate business people with a broad understanding of the world around them, of what forces keep the wheels of business turning and of the measures needed to advance financial reporting to more informative and useful levels. Meeting the competition takes more than technical proficiency.

**RULE-BASED STANDARDS**

On the surface, the sheer bulk of accounting and auditing standards — to say nothing of the wealth of regulations emanating from assorted government agencies — lends credence to the proposition that ability to read the rulebook has become the primary professional qualification.

It may look that way, but it just isn’t so. There is no way standards setters and regulators following a cookbook philosophy can cover all the bases. They cannot hope to offer prescriptions for dealing with types of transactions that have yet to be invented. They cannot hope
to stay in front of the fertile minds bent on finding loopholes in existing standards — new or old.

Consider some of the arcane transactions under consideration by the FASB and its new Emerging Issues Task Force: collateralized mortgage obligations, convertible debt sweeteners, Ginnie Mae dollar rolls, termination of interest rate swaps, stock option pyramiding. And there will be more in a world that found ways to end run the FASB's ill-fated leasing standard and discovered "instant" in-substance defeasance while the Board's "Extinguishment of Debt" standard was still rolling off FASB presses.

Paradoxically, it takes equal, if not more, skill to practice in a rule-based environment than in one where basic principles and objectives are to be applied. The choices when a transaction does not quite fit existing standards are apt to be multiple rather than singular, adding to the difficulty of identifying the proper accounting for a discrete set of facts and circumstances.

It takes considerable skill to convince a reluctant client that an apparently promising loophole is a lump of lead, not a gold bar. Confrontation — standing firmly on the rulebook — will not work. Individuals who do not listen, who do not reason with clients will not prevail. Important in any environment, the skill of persuasion becomes even more essential when the complexities of a rule-based system multiply pitfalls and opportunities for avoidance.

You as educators are squeezed between the rising tide of rules and the fixed parameters of your educational programs. More time devoted to the rulebook's complexities means less time for instruction in accounting concepts and in how to apply them, less time for exploring the nonaccounting facets of business and the fascinating world in which it operates, less time to develop vital communicative skills, less time to acquire the ability to reason.

Something has to give. I am convinced that balance is primary; preoccupation with the rulebook must not be permitted to crowd out the essentials of a broad-based education. Professionals need to know the scope of the rulebook, how to recognize its pitfalls and imperfections and how to apply it. They do not need to memorize its contents. Those equipped with the profession's nontechnical qualifications should have no difficulty in extending the scope of their technical knowledge without the prod of a formal program.

**AUDIT METHODOLOGY**

If auditing was slow to enter the computer age, competition provided the spur to hasten the process. Two effects stand out.
First, nearly total reliance on computerized financial records has aroused client awareness of the dangers lurking in the invisible depths of computer systems and has heightened client expectations of auditors. Auditors are expected to understand the full parameters of the client's financial systems and to provide assurance that all is well or, alternatively, to present positive recommendations on what is needed to make the system safe and effective. If we are to be responsive to service needs, we must meet those expectations — with obvious implications for our people requirements.

Second, the computer is replacing the legions of inexperienced staff assistants that once labored to gather audit evidence. The emphasis has swung sharply to planning, analysis and investigation of expectations generated with computer assistance. Always essential, knowledge of how business operates, how management controls it and how the external environment affects it has become even more critical.

Coordination with internal audit groups — whose professional stature and competence has risen dramatically in the past several years — has assumed increased importance. Audit decision-making requiring the attention of senior engagement management has become a nearly continuous process.

The route from raw recruit to seasoned professional was once a long, slow process. Professionals were molded over time — time to instill professional values — independence of mental attitude, objectivity and ethics — time to develop the ability to make sound audit decisions and accounting judgments, time to acquire a fine-tuned understanding and awareness of the multifaceted economic, political, and social forces that often determine business success or failure.

That seasoning time has become a vanished luxury, severely truncated by changes in audit methodology. The qualifications once expected halfway up the ladder are now desperately needed on the bottom rung. The hurning question is how to bridge the gap. Is there a substitute for seasoning the caldron of experience?

Professional firms commit enormous resources to continuing education programs — programs extending throughout the span of a professional's career. We are conforming these programs to today's requirements, attempting to bring new entrants up to speed at a faster clip. Strengthened supervisory systems also help bridge the gap.

But it is clear to me that more is needed. Quite simply, today's entrants must be better equipped when they join our ranks — better equipped in the nontechnical areas I have been referring to. The
methods of achieving that goal lie squarely within your field of expertise.

AUDIT/NONAUDIT SERVICES

The changing composition of professional services — the last of the environmental conditions identified in my opening remarks — relates more to professional opportunities than to professional qualifications. As professional firms offer a broader range of services, tax and management advisory services are grabbing an ascending share of the total professional pie. Through the function of simple arithmetic, audit services are seen as declining in importance.

The provision of services that today would be classified as "non-audit" began very early in the profession's history. The recent surge in nonaudit services is rooted in our profession's response to the business world's demand for assistance in coping, for example, with technological advances, with international expansion plans, with the ever more intricate requirements of taxing jurisdictions — federal, state and local, national and international. Supplying such services means increased opportunities for professionals interested in the numerous branches of the consulting field and for those attracted to careers in taxation.

But what of auditing? Analysts concentrating on the profession would have us believe that auditing has limited potential. They paint a gloomy picture of professional firms scrambling for position in a market with fixed — if not shrinking — margins. Does this mean that the attest function should be avoided as a career opportunity?

Not at all! The apparent decline is a mirage born of improved audit efficiencies, translated into fewer hours per engagement, and the expansion of nonaudit services.

Improved efficiency does not mean a dying art of static demand. Absent the unlikely abandonment of federal requirements that the financial statements of public companies be audited, auditing is here to stay. Further, the attest function has never reached its full potential in sectors where no statutory requirements exist. Governmental activities, for example, present a potential market far from saturation.

The demand for financial audits is by no means fixed. The face of the business world is constantly shifting as a result of mergers, divestitures, and changes in direction. Older businesses mature and new ones emerge; world population increases and economies grow. The demand for professional services in all categories follows.

As counselors to your students, you should recognize that the
modernized brand of auditing has not lost its opportunities. Indeed, with much of the early drudgery stripped away, it has become more exciting and more rewarding for those motivated and qualified to accept its challenges.

CONCLUSION

The pressures imposed by our practice environment are molding our profession as never before. Heightened public expectations, intense competition, proliferating technical rules, and audit methodology designed for this computer age, each in its own way, have left permanent imprints on professional practice. The broader range of services offered by professional firms is expanding the opportunities available in nonaudit disciplines. And change will continue.

I firmly believe that a considered response to environmental demands will leave the profession stronger and further raise its stature.

But to do so, we need the people resources to match the demands on the profession. Although the ante is rising in that area also, technical proficiency is just the beginning — a prerequisite accompanied by polished communicative and interpersonal skills.

The profession needs leaders not followers. It needs pioneers and entrepreneurs — problem solvers and innovators who focus on solutions and ask “Why not?” rather than guardians of yesterday who shrink from problems and ask “Why?” It needs interpreters of the spirit of informed financial reporting not recipe readers. Above all, it needs people with a broad-based view not only of what makes business tick but of the world around them, people who are with the times and on the cutting edge of techniques and technology.

Where do we find them? The answer in large part lies in your bailiwick — in academic institutions, like this one, responsive to the professional climate its graduates will face. It’s a tall order, but I am confident you are equal to the challenge.
Audit (sampling) risks can be quantified and controlled when statistical procedures are employed. Consequently, statistical sampling has become an important tool for auditing financial statements. The purposes of this paper are to (1) describe our understanding of the role of statistical sampling in auditing, (2) discuss the difficulties of application arising from the "rare population" phenomenon, and (3) introduce model-based sampling as a means of overcoming these difficulties.

Our presentation is organized into five sections. The first describes the problems of applying statistical sampling methods to auditing populations and introduces the concept of rare populations. The second section reviews two methods that currently are proposed and used to overcome problems attendant with rare populations. Model-based sampling is presented, in the third section, as a more general approach to addressing these problems. The fourth section reviews the commonly used ratio model in an auditing context both from

Editors' Note: This paper was anonymously refereed prior to acceptance.
the perspective of design-based and model-based sampling. Research implications of the foregoing are presented in the final section.1

PROBLEMS OF APPLYING STATISTICAL SAMPLING IN AUDITING

A statistical sampling strategy consists of two distinct procedures. The first procedure is sample selection or design which involves how a sample is chosen. For example, random sampling could be employed by which the sampler assigns a selection probability to each item in the population. Alternatively, the sampler could use nonrandom sampling and, thus, choose specific items in the sample in some purposeful manner. The second distinct component of a sampling strategy involves the method used to project the sample evidence to population parameters. Conventional sampling strategies rely on random sampling as a basis for such inference. However, characteristics of accounting populations complicate the application of these sampling methods and lead to poor performance of the commonly used estimators with the sample sizes typically employed in auditing (that is, less than 250). These problematic characteristics include low error rates and high error distribution skewness.

Background and Source of the Problems

The often-cited poor performance of statistical sampling in auditing is due to actual audit value confidence intervals being narrower than nominal intervals. That is, the actual confidence level obtained from a sample deviates significantly from the nominal confidence level based on conventional sampling theory for economically viable sample sizes. Consequently, the auditor incurs additional risk that is not explicitly measured. Substantial research has demonstrated that the actual risk level often incurred is significantly higher than the nominal level that the auditor planned to accept.

Computer simulations have been the most common and powerful tool in many of these research studies (Kaplan 1973; Neter and Loebbecke 1975; Reneau 1978; and many others). Stringer (1963) referred to the possibility of poor performance when error rates are low (that is, rare). Anderson and Teitlebaum (1973) pointed out similar problems with both attribute and variable sampling methods. Johnson, Leitch, and Neter (1981) and Ramage, Krieger, and Spero (1979) provided evidence that accounting populations have low error rates and are highly skewed. These studies provide the empirical
basis of an explanation for the problems auditors have encountered in practice.

What is the statistical basis of this poor performance? Suggested sources are (1) correlation of the estimate of the mean with the estimate of the standard deviation, (2) understatement of the estimate of the standard deviation, and (3) bias in the mean estimate. These means and standard deviations refer to estimates of audit values but also could apply to estimates of errors with error defined as book value minus audit value.

Kaplan (1973) reported an association between the estimate of the mean audit value and the estimated standard deviation. Any such correlation would violate assumptions of independence of these two estimates and lead to poor performance. Because of this problem, Kaplan questioned the appropriateness of using the t-distribution.

Stringer (1963) and Anderson and Teitlebaum (1973) pointed to low error rates as the major source of the problem. They indicated that when no errors are found in the sample and difference or ratio estimators are used, the estimate of the standard deviation is zero. This implies no uncertainty in estimation, which is impossible without making a census of the population. Due to this logical fallacy in the circumstance of no sample errors, these authors suggested the possibility of substantial underestimation of the standard deviation as a source of the poor performance. Such underestimation is clear when the sample contains no errors. However, the expected value of the standard deviation estimate is not necessarily understated in this circumstance since expectation is taken with respect to repeated samples. That is, the standard deviation computed from a particular sample may underestimate the population standard deviation, but the sample standard deviation still may be an unbiased estimator for the population standard deviation.

Duke, Neter, and Leitch (1981) suggested that the poor performance was a consequence of the lack of normality of the unit normal deviate distribution due to small sample sizes. They constructed distributions of several versions of the standardized unit normal deviate and tested these for normality. For example, to correct for the possible effects of bias in the standard deviation estimate, the population standard deviation (which was known because of the simulation) was substituted. Also, to correct for the possible effects of dependence of the estimate of the mean with the estimate of the standard deviation, standardized unit normal deviates were constructed with randomized pairings of the two estimates. They concluded that no bias of the estimated standard deviation, bias of the
point estimate, or dependence of the estimate of the mean and the
estimate of the standard deviation fully explain the lack of normality
of the unit normal deviate distribution.

Approximate Normal Distribution

Conventional sampling strategies rely heavily on the approximate
normal distribution of the sample estimate to provide reliable infer-
ence when the sample size is large. For simple random sampling with
the mean-per-unit estimator, the Central Limit Theorem states that
for any population with finite variance, the distribution of sample
means is normal when the sample size is sufficiently large. A confidence
interval, thus, can be constructed for the population value of variables
under consideration. Such approximate normality for large samples
also is the basis for inferences made by auditors when the ratio and
regression estimators are employed.

However, the sample size needed to achieve the approximate
normality of the estimate depends on the population characteristics
of variables under investigation. In auditing one is concerned about
the magnitude of accounting error (that is, the difference between
the book value and the audit value of accounts). As indicated by
Ramage, Krieger, and Spero (1979) and Johnson, Leitch, and Neter
(1981), populations of accounts receivable tend to have low error
rates. In addition, the error amount distributions of most populations
typically are characterized by high peaks around the mean, positive
skewness, and fat upper tails. Accounting error populations, therefore,
generally, are far from normal. Such populations would require a
very large sample size for the distributions of the sample estimates
of population errors or audit values to be approximately normal,
and, thus, permit reliable confidence intervals to be constructed.

The poor performance of classical sampling strategies on auditing
populations is not due to any failure of conventional sampling theory.
In fact, given the common characteristics of auditing populations,
conventional sampling theory predicts poor performance for the
sample sizes commonly used by auditors. Larger sample sizes are,
however, too costly for the objectives of the financial audit. Thus,
auditors should not be surprised when small sample sizes and sampling
strategies fail to achieve the nominal confidence level that should
have resulted from large sample normal approximation theory.

Rare Populations

The auditor often deals with populations in which the event of
interest occurs only rarely. Cox and Snell (1979) discuss various
solutions to the failure of classical statistical methods when used in connection with rare populations, including a Bayesian solution, when errors are of the same sign and a maximum fractional error is known.

Audit sampling also is complicated because accounting errors possess characteristics of both attributes and variables (Anderson and Teitlebaum 1973; and Goodfellow, Loebecke, and Neter 1974a and 1974b) and both need to be considered. Smith (1976, 1979) states that the only available sampling theory for rare populations is attribute sampling theory and that "a good theory for rare events that are values, not attributes, is badly needed."

In summary, because of the rare population phenomenon, sample sizes required to justify reliance on normal approximation in auditing applications are very large — so large that in many cases auditors have made the judgment that the cost of sampling is not economically justified. Auditors have currently turned to two methods of addressing this problem which are discussed in the next section.

TWO METHODS FOR APPROACHING THE RARE POPULATION PROBLEM

To circumvent the difficulties associated with the rare population problem, adaptations of classical sampling strategies have been proposed along two basic lines. One such adaptation is to compensate for the coverage of the nominal confidence limits by modifying either the sample standard deviation (Jones 1975; Frost and Tamura 1982) or by modifying the multiple of the estimated standard deviation used in computing the usual large sample confidence limits (Garstka and Ohlson 1979). A second approach is to compute direct estimates of an upper bound on total error by developing sampling strategies that are not based on large sample confidence limits (Stringer 1963; Anderson and Teitlebaum 1973; Fienberg, Neter, and Leitch 1977; Godfrey and Neter 1982; and Dworin and Grimlund 1984).

Approach One — Compensating for Coverage

Jones (1975) proposed augmentation of the sample standard deviation, in those cases in which few errors are found in the sample. The amount of augmentation is based on an intuitively derived estimate for the potential standard deviation understatement. Frost and Tamura (1982) used the jackknifed ratio estimator (with little success) to improve the estimate of variance in a variety of accounting populations. Similarly, Garstka and Ohlson (1979) proposed a modification of the multiple of the estimated standard error. Specifically,
they suggested a multiple based on both the sample size and the number of nonzero differences found in the sample. Such a multiple would be larger than the unit-normal deviate, but for large sample sizes and large numbers of nonzero differences, it would approach the unit-normal deviate.

**Approach Two — Direct Estimation of an Upper Bound**

Stringer (1963) proposed a method for obtaining a direct estimate of the upper bound that did not rely on large sample theory. Teitlebaum (1973) investigated some of the theoretical properties of the Stringer bound, while Anderson and Teitlebaum (1973) proposed the use of dollar-unit sampling in conjunction with the Stringer bound. Goodfellow, Loebbecke, and Neter (1974a and 1974b) offer a critical review of this procedure.

Fienberg, Neter, and Leitch (1977) proposed a different method for direct estimation of an upper bound based on the multinomial distribution. This method is computationally intensive and involves nonlinear optimization techniques. Testing by computer simulation of this method is reported in Neter, Leitch, and Fienberg (1978) and Leitch et al. (1982).

Godfrey and Neter (1982) proposed Bayesian bounds for dollar-unit sampling. Their work is an extension and test of the Cox and Snell (1979) model which relies on explicit assumptions about prior distributions on error rate and error tainting to overcome the difficulties resulting from the rare-error phenomenon. They found that conservative priors will assure adequate coverage for a wide range of populations with greater efficiency than the Stringer bound.

Dworrin and Grimlund (1984) proposed a bound based upon assumed distributions of error rate and observed individual errors (error taintings) to obtain an upper confidence bound for the mean error. This bound uses the first three moments of the error distributions and, hence, explicitly considers the skewness of the sample elements. They reported that this moment bound performs well in populations with both overstatement and understatement errors. Summaries of the two approaches are provided in Tables 1 and 2, respectively.

The two approaches to the rare population problem have provided sampling methods that explicitly address its manifestations (for example, low error rates and high skewness of error amounts or taintings). From the foregoing discussions and summaries in Tables 1 and 2, it is clear that accounting researchers (especially those using the second approach) are moving toward modifying conventional
Table 1. Summary of Proposed Methods to Deal with Rare Error Problem under Approach One — Compensating for Coverage

<table>
<thead>
<tr>
<th>Proposed Method</th>
<th>Approach Taken</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Augmentation of sample standard deviation (Jones 1975).</td>
<td>Use of an intuitively derived statistical estimator.</td>
<td>Some improvement on reliability, but is uncertain beyond the simulated populations.</td>
</tr>
<tr>
<td>2. Modified ratio estimator with pps sampling design (Garska and Ohlson 1977).</td>
<td>Use of C-statistics in lieu of t-statistics and z-statistics where C-statistics depends on both the sample size and the numbers of errors in the sample.</td>
<td>Confidence interval under C-statistics is wider than that under t or z statistics and is closer to the nominal levels.</td>
</tr>
<tr>
<td>3. Jackknifed ratio estimator (Frost and Tamura 1982).</td>
<td>Use of sampling splitting to obtain estimate of variance.</td>
<td>Unable to overcome the rare error problem but provide more reliable confidence interval than conventional ratio estimator.</td>
</tr>
</tbody>
</table>

sampling methods in the inferential process by formally incorporating assumptions about the population characteristics of accounting errors. Simulations (Duke, Neter, and Leitch 1982; and Reneau 1978) indicate that one universally best procedure may not exist, which also is the conclusion reached by statistical theory (Smith 1976). Different sampling strategies perform well on different types of populations.

Each of these strategies continues to rely on conventional design-based sampling concepts. In design-based sampling, some form of a random sample is to be drawn from the population of interest. Estimates of population parameters then are derived based on the random sample selection methodology. However, the rare occurrence problem generally results in an unstable estimate of variance and the resulting failure of the estimator.

We began our study of an alternative approach, model-based sampling, with the expectation that it might ameliorate the rare population problem. In model-based sampling, the sampler uses prior information concerning the likely form of the population and derives optimal sampling strategies generally of a nonrandom nature based on this information. The next section describes model-based sampling concepts and compares them with the classical design-based concepts described above.
Table 2. Summary of Proposed Methods to Deal with Rare Error Problem under Approach Two — Direct Estimation of Upper Error Bound with DUS Sampling

<table>
<thead>
<tr>
<th>Proposed Method</th>
<th>Approach Taken</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Stringer Bound (Stringer 1963; Anderson and Titlebaum 1973).</td>
<td>Use binomial or Poisson distribution on observed error in the sample to estimate population error rates. These error rates are converted to error amount with some assumptions of error taintings.</td>
<td>Providing upper error bound even if no error is found in the sample. But the upper bound is too conservative (wide).</td>
</tr>
<tr>
<td>2. Multinomial Bound (Fienberg, Neter, and Leitch 1977; Neter, Leitch, and Fienberg 1978).</td>
<td>Assuming the occurrence of error has multinomial distribution with 101 possible outcomes in cents.</td>
<td>The upper error bound is generally tighter than Stringer bound. But it is computation-intensive especially for larger number of errors.</td>
</tr>
<tr>
<td>3. Bayesian Bound (Cox and Snell 1979; Godfrey and Neter 1982).</td>
<td>Assuming the unconditional error probability and the reciprocal of the mean tainting follow gamma prior distributions. The number of dollar units in error in the sample is assumed to have Poisson distribution. The tainting in the sample is assumed to have an exponential distribution. The posterior distribution of mean tainting in the population can be obtained and used to compute the upper error bound.</td>
<td>This bound is highly efficient when error amounts are low and moderately efficient for populations with large error amounts.</td>
</tr>
</tbody>
</table>

MODEL-BASED SAMPLING

There exist two basic competing philosophies in the theory of sampling inference from finite populations, design-based and model-
Table 2. Continued

<table>
<thead>
<tr>
<th>Proposed Method</th>
<th>Approach Taken</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>4. Moment Bound (Dworin and Grimlund 1984)</td>
<td>Combination of the first three moments of error rate (assuming beta distribution) and error tainting leads to the distribution of mean error. It is represented by a three-parameter gamma distribution. Confidence interval is then approximated from this distribution.</td>
<td>This bound is much tighter than Stringer bound. Under some circumstances, it is also tighter than multinomial bound.</td>
</tr>
</tbody>
</table>

based. Classical survey sampling (as discussed above, presented by Neyman (1934) and covered by most traditional sampling texts such as Cochran (1977)), is design-based. In design-based sampling, the sampler does not have to make assumptions about the population distributions to support an inference. Rather, the sampler estimates the population total of some variables from the result of a random sample. Probabilistic inferences can be made because of the randomness induced by the sample design. Reliance is placed on the approximate normal distribution of sample estimates of population parameters.

Model-based sampling theory proposes the existence of a superpopulation of random variables from which the actual population units are realized outcomes. The superpopulation is modeled to reflect available knowledge of the variables, and both the sample drawing procedure and the inference procedure may be derived from the model of the superpopulation. In model-based sampling, the sampler does make assumptions about the population distributions, but does not have to rely on random sample selection procedures. The sampler predicts the values of those items not included in the sample from the sample result. Probabilistic inferences can be made from nonrandom samples because the randomness (probability structure) is embedded in the assumed superpopulation model. This point will be further discussed below. Chart 1 contrasts the general difference between design-based and model-based sampling.

In auditing contexts, one may assume that there exists a superpopulation for the generation of errors that may enter a client’s records.
Chart 1. Contrast between Design-based and Model-based Sampling

Design-based Sampling Approach


design-based sampling approach

Model-based Sampling Approach

model-based sampling approach

Superpopulation Model

Possible random manifestations

Nonrandom representative sample

(Actual population)
This superpopulation encompasses all possible errors (with perhaps unequal probabilities) over all firms, all accounts, and all time. One also may assume that there are several types of filters in a company, which are the systems employed to limit the kinds of errors or modify the probabilities of their occurrence in a specific set of accounting records. Two such filters are the company’s (1) specific accounting system and (2) system of internal controls.

The accounting system chosen by the company determines the kind of errors that may enter the accounting records. Internal control systems are designed to detect certain forms of error and are biased in their detection. For example, if the company has a computer-based accounting system, the possible errors and their probabilities would be different from a manual system. Also, the errors in an inventory system are different from those in accounts receivable or long-term liabilities.

The auditor examines these two filters because they affect the possible errors that are in the accounts. Generally accepted auditing standards require that auditors conduct a systems review and an internal control review to gather evidence about possible errors and the likelihood of their occurrences and to serve as a basis for audit planning.

Thus, there are two superpopulation models, a global model, and one that has been filtered by the accounting system and the system of internal controls. Auditors have general knowledge of the global model and they gather specific knowledge during the audit as to how errors may be modified by filters. It is the filtered superpopulation that is of interest in applying model-based sampling strategies.

While model-based sampling derives its inferential ability from a superpopulation model, design-based sampling bases its inference on the sample design procedure. In conventional design-based sampling strategies, the repeated large sample estimates of the population mean are approximately normally distributed. With this normal approximation, we can construct confidence intervals for the population mean with specified reliability. In the first approach used by accounting researchers to improve the performance of the conventional sampling strategies, the random sample design is employed with modification of the standard deviation estimate or its multiple. In the second approach, the direct estimate of the upper-error bounds by dollar-unit sampling strategies, the sampling unit is the individual dollar. The auditor applies a simple random sampling design procedure on individual dollar units and makes certain assumptions about the distributions of error occurrences and taitings for inference. Basi-
cally, these two approaches continue to apply random-sample design as a basis for their inference and thus can be considered as design-based sampling.

Model-based theory is a third alternative. The auditor would model the superpopulation based on his extensive prior information about accounting populations. Sample drawing procedures and sample inference procedures then would be derived from this model, thus allowing the auditor to tailor the sample plan to the modeled characteristics of the population.

**Model-based Sampling**

When model-based sampling is used in auditing, the actual population of audit values, denoted by \( y = (y_1, y_2, \ldots, y_N) \) is a realized outcome of the random vector \( Y = (Y_1, Y_2, \ldots, Y_N) \) which has an \( N \)-dimensional joint distribution, also referred to as the superpopulation or the superpopulation distribution. The auditor’s prior information is used to model the superpopulation either in a simple or detailed fashion. A simple model could involve specification of the first two moments of \( Y \). A more complex model could be based upon the assumption that \( Y \) is some linear function of an auxiliary variable \( x \) which could be the book value in an audit application.

Given an assumed superpopulation model, the next step is to select the sampling strategy which consists of a sample design (or selection) procedure (denoted by \( p \)), and an estimator (denoted by \( T \)). The sampling strategy chosen should be optimal according to objective criteria. Sarndal (1978) indicated six possible objective criteria for the selection of an optimal sampling strategy. These criteria are in the form of minimizing the mean square errors (MSE) of the estimators with respect to the probability structures induced in the sample design procedure and/or assumed in the superpopulation models.

When there exists an auxiliary variable (for example, the account book value, \( x \)) that is closely related to the primary variable of interest (the audit value, \( y \)), a superpopulation model for \( y \) can be constructed: for example:

\[
Y_i = \beta x_i + \epsilon_i(x)^\gamma, \text{ for } i = 1, 2, \ldots, N,
\]

where \( \epsilon_i \) is the statistical error term and the expected value of \( \epsilon_i \):

\[ E(\epsilon_i) = 0, \]

the variance of \( \epsilon_i \):

\[ V(\epsilon_i) = \sigma^2, \]

an unknown constant, and the covariance of \( \epsilon_i \) and \( \epsilon_j \) with \( i \neq j \):
\[ \text{COV}(e, e') = 0. \]

The foregoing assumes a regression through the origin and the ratio model. If the criterion chosen is to minimize the variance of the population mean estimate, the best linear unbiased (BLU) estimator would be the ratio estimator and the optimal sample would be that consisting of the largest \( x_i \) from the population for a given size \( n \). This result is intuitively appealing but is contingent upon the assumption that the regression line passes through the origin. This point is developed in the next section of the paper.

**Comparison of Design-based with Model-based Sampling**

Smith (1976), Cassel, Sarndal, and Wretman (1977), and Sarndal (1978) have examined both design-based and model-based sampling in detail. In general, classical statistics on nonfinite populations assume that the functional form of the population is known and that inference can be based on this knowledge. On the other hand, design-based inference for finite populations does not require knowledge of the population distribution. Given an adequate sample size and using proper sample selection design (simple random sampling, stratified random sampling, cluster sampling, probability-proportional-to-size sampling, and so on), the Central Limit Theorem allows the sampler to assume that the distribution of the estimator is approximately normal. In practical auditing applications, sample sizes that are economically small are not adequate, however, to permit reliance on the Central Limit Theorem.

Generally, auditors know a great deal about the distribution of the accounting population. They know with a high degree of certainty, for example, the distribution of population book values. Additionally, they know that audit values are highly correlated with book values. They also have knowledge about the types and probabilities of various forms of accounting errors in the population both from prior audits and from their evaluation of internal controls.

Model-based sampling allows the auditor to utilize his/her substantial knowledge of accounting population distributions. Moreover, use of model-based concepts may allow auditors to circumvent the sample size problem encountered in design-based approaches. In design-based sampling, the rare population phenomenon leads to unstable variance estimation problems with the relatively small sample sizes used by auditors. In model-based sampling, knowledge of the error-generating process provides information on which an appro-
appropriate superpopulation model might be constructed and reliable strategies might be derived.

**Protection against Misstatement of the Model**

A crucial question related to model-based sampling is "What happens if the model is wrong?" Such model misstatement has been referred to as the robustness of the sampling strategy (Royall and Herson 1973a). Balanced sampling is a method of achieving robustness in model-based sampling in which the sampling unit is selected according to specified criteria. As defined by Royall and Herson (1973a), a sample is balanced if for any level of power, the mean of the sample auxiliary variable raised to that power equals the mean of the population auxiliary variable also raised to that power. That is,

\[ \bar{x}_j^{(\beta)} = x^{(\beta)} , \]

(2)

where

\[ \bar{x}_j^{(\beta)} = \frac{1}{n_j} \sum x_j^i \text{ and,} \]

\[ x^{(\beta)} = \frac{1}{N} \sum x_j^i , \]

for \( j = 0, 1, \ldots, J \).

Balanced sampling is neither random nor judgmental. Rather, it is a purposive sample selection procedure with a specific optimization goal. In addition, the auditor may stratify the population based on book values, and then apply balanced sampling and the ratio estimator within each stratum (Royall and Herson 1973b).

Consider the case in which the auditor obtained a sample and the mean book value of that sample was far from the mean book value of the population. How much comfort could the auditor take in any inferences made from that sample? Such a sample could be considered nonrepresentative. We observe a recognition of this idea in design-based systems when an adjustment for the sample is made in difference and ratio estimators. An example of a model-based sampling strategy, having intuitive appeal in auditing, is a selection of a nonrandom sample of a given size for which the sample’s mean book value most closely equals the known population’s mean book value. Criteria for balancing samples may be more complex, including fit requirements for higher moments of the sample which tend to be computationally intense. However, algorithms are available to reduce this computational burden.\(^4\)
In summary, model-based sampling strategies involve determining an appropriate superpopulation model and deriving robust sample selection methods and estimators. The sample selection methods do not have to be randomization-based, but should be able to deal with a variety of possible superpopulation models. Design-based strategies based on random sampling require large sample sizes to deal effectively with typical accounting populations. Model-based strategies make use of the auditor's prior information to derive sampling strategies to achieve efficiency, but must include protection against model misspecification. Balanced sampling offers one such protection. In the following section, we examine a special superpopulation model, the ratio model. In the process, we obtain further insights for the operationalization of model-based sampling and demonstrate their potential implications in the audit environment.

**BALANCED SAMPLING AND THE RATIO MODEL**

Many issues have to be addressed to apply a model-based approach appropriately in the audit context. The issues to be addressed in this section include (1) if we are certain as to the superpopulation model, what is the best linear unbiased (BLU) estimator and what is the optimal sample selection design? and (2) if the assumed model is not known with certainty to be appropriate, can we still use the sampling strategies derived from question (1) or do we have to use a different sampling strategy to obtain unbiased and efficient estimates?

To exemplify the approach toward general solutions, we first consider a superpopulation model based on a linear regression line through the origin — a ratio model. A ratio model is assumed because audit values usually are very close to the account book values. The auditor estimates the total audit value and constructs a confidence interval at specified reliability. A comparison between confidence bounds and the materiality level will give evidence as to the existence of any material error.

A ratio model that relates the audit value to book value is presented. The auditing implications are adapted from the theoretical development by Royall (1979) and Royall and Herson (1973a and 1973b). The theorems are provided by Royall and Herson (1973a). Notations to be used for further discussions are listed in Chart 2.

**Ratio Model**

The model to be examined is of the following form:

\[
Y_i = \beta x_i + \epsilon_i, \tag{3}
\]
Chart 2. Summary of Notations

\[ \sum_i \quad \text{Summation over population items} \]

\[ \sum_s \quad \text{Summation over the items in the sample s} \]

\[ \sum_t \quad \text{Summation over the items not in the sample} \]

\{[\delta_0, \delta_1, \ldots, \delta_J; g(x)] \} \quad \text{Represents the model:} \]

\[ y_i = \delta_0 \beta_0 + \delta_1 \beta_1 x_i + \delta_2 \beta_2 x_i^2 + \ldots + \delta_J \beta_J x_i^J + \epsilon_i (g(x)) \]

for \( i = 1, 2, \ldots, N \)

where \( \delta_j \) are indicator variables, with values 0, or 1

for \( j = 1, 2, \ldots, J \)

\( \epsilon_i \) is the error term

\( g(x) \) is a function of \( x_i \)

\[ \hat{\theta}[\delta_0, \delta_1, \ldots, \delta_J; g(x)] \quad \text{The BLU estimator under the model} \]

\{[\delta_0, \delta_1, \ldots, \delta_J; g(x)] \}

where \( Y_i \) = primary variable of interest, for example, audit value;

\( x_i \) = auxiliary variable, for example, book value;

\( \epsilon_i \) = the random-error term; and

\( E(\epsilon_i) = 0 \)

\( V(\epsilon_i) = \sigma^2 \)

\( \text{COV}(\epsilon_i, \epsilon_j) = 0 \), for \( i \neq j \).

In this model, the audit value \( Y_i \) is a linear function of the book value \( x_i \) with some random error \( \epsilon_i \). This random error is assumed to have a zero expectation and a constant variance \( \sigma^2 \). Additionally, the covariance of the errors of any two different items is zero. The constant error term variance also implies that the variance of the audit values is proportional to the book value. It is noteworthy that the probability structure is embedded in the specification of the error term \( \epsilon_i \). However, as was noted earlier and will be demonstrated later, the implication for sample selection procedures deriving from the
superpopulation model differs from those commonly employed by auditors who use the more familiar design-based ratio estimators.

Theorem 1. The ratio estimator $\hat{T} = (\hat{y}_i / \hat{x}_i) \sum x_i$ is the best linear unbiased (BLU) estimator for $T = \sum y_i$ under Model (3). Its variance,

$$V(\hat{T}) = \sigma^2 (\sum \frac{x_i}{\hat{x}_i} - n\hat{x}_i) \sum x_i / n\hat{x}_i.$$ 

Proof: (a) BLU estimator for $T$:

(4) Let $d_i = e_i(x_i)\hat{x}_i$.
Then, $E(d_i) = 0$.
and $V(d_i) = \sigma^2 x_i$.

(5) Let $T = \sum \hat{y}_i = \beta \sum x_i + \sum d_i$;

(6) then, $E(T) = \beta \sum x_i$.

(7) Let $\hat{T} = \sum a_i \hat{y}_i = \beta \sum a x_i + \sum a d_i$;

(8) then, $E(\hat{T}) = \beta \sum a x_i + \sum a E(d_i) = \beta \sum a x_i$.

(9) $V(\hat{T}) = V(\sum a x_i) + V(\sum a d_i) = \sigma^2 \sum a x_i$.

From (5) and (8), it can be seen that $\hat{T}$ is model-unbiased if $\sum a x_i = \sum x_i$. To obtain the minimum variance for $\hat{T}$, we shall minimize $V(\hat{T}) = \sigma^2 \sum a x_i$, subject to $\sum a x_i = \sum x_i$.

Let $r = $ Lagrange multiplier.

(10) Then, $z = \sigma^2 \sum a x_i + r(\sum x_i - \sum a x_i)$.
Let $\partial z / \partial a_i = 2\sigma^2 a x_i - r x_i = 0$; and
$\partial z / \partial r = \sum x_i - \sum a x_i = 0$.
From (10), $2\sigma^2 a x_i = r x_i$,
and $a_i = k$ (constant) = $(1 / n\hat{x}_i) \sum x_i$. 
\( k \) must be equal to \( \left( 1/n\hat{x} \right) \sum_1^N x_i \) so that the condition for the unbiasedness, \( \sum_j a_j x_i = \sum_1^N x_i \), can be satisfied. Therefore, the BLU estimator for \( \bar{T} \) is

\[
\hat{T} = \sum_j a_j y_j \\
= \sum_j \left( \left( 1/n\hat{x} \right) \sum_1^N x_i \right) y_j \\
= \left( 1/n\hat{x} \right) \sum_1^N x_i \sum_j y_j \\
= \left( y_i / \hat{x} \right) \sum_1^N x_i
\]

This \( \hat{T} \) is the ratio estimator, which also is the BLU estimator under the model in (3).

(b) Variance of Ratio Estimator: As \( \hat{T} = \beta \sum_1^N x_i + \sum_1^N d_i \):

\[
\hat{T} = \sum_j a_j y_j = \beta \sum_j a_j x_i + \sum_1^N a_j d_i
\]

\[
\hat{T} - T = \left( \beta \sum_j a_j x_i + \sum_1^N a_j d_i \right) - \left( \beta \sum_1^N x_i + \sum_1^N d_i \right)
\]

\[
= \beta \left[ \sum_1^N x_i / n\hat{x} \sum_1^N x_i + \sum_1^N (x_i / n\hat{x}) d_i \right] - \left( \beta \sum_1^N x_i + \sum_1^N d_i \right)
\]

\[
= \beta \left[ \sum_1^N x_i / n\hat{x} \sum_1^N x_i + \left( \sum_1^N x_i / n\hat{x} \right) \sum_1^N d_i \right] - \left( \beta \sum_1^N x_i + \sum_1^N d_i \right)
\]

\[
= \left( \sum_1^N x_i / n\hat{x} \right) \sum_1^N d_i - \left( \sum_1^N d_i + \sum_1^N d_i \right)
\]

\[
= \left( \sum_1^N x_i / n\hat{x} - n\hat{x} \right) \sum_1^N d_i - \sum_1^N d_i
\]

Then, \( V(\hat{T}) = \mathbb{E}[(\hat{T} - T)^2] \):

\[
\mathbb{E}\left[ \left( \sum_1^N x_i / n\hat{x} - n\hat{x} \right) \sum_1^N d_i - \sum_1^N d_i \right]^2
\]

\[
= \mathbb{E}\left[ \left( \sum_1^N x_i / n\hat{x} \right)^2 - \left( n\hat{x} \right)^2 \sum_1^N d_i \right]^2 + \mathbb{E}\left( \sum_1^N d_i \right)^2
\]

\[
= \left[ \left( \sum_1^N x_i / n\hat{x} \right)^2 \sum_1^N \mathbb{E}(d_i)^2 + \sum_1^N \mathbb{E}(d_i)^2 \right]
\]

\[
= \left[ \left( \sum_1^N x_i / n\hat{x} \right)^2 \sigma^2 \sum_1^N x_i + \sigma^2 \sum_1^N x_i \right]
\]
\[
\pi = \sigma^2 \left( \sum_{i=1}^{N} x_i - n \bar{x}_{1n} \right) / \left( n \bar{x}_{1n} \right) + \sigma^2 \left( \sum_{i=1}^{N} x_i - n \bar{x} \right)
\]

\[
= \sigma^2 \left( \sum_{1}^{N} x_i - n \bar{x}_{1n} \right) \sum_{1}^{N} x_i / n \bar{x}_{1n}
\]

(11)

From (11), one can see that to reduce the variance of the estimator, \( V(\hat{\pi}) \), either the denominator would have to be increased or the numerator decreased. Since \( \sigma^2 \) is a fixed but unknown quantity, \( \sum_{1}^{N} x_i \) is a fixed and known quantity, the only item left is \( \bar{x}_{1n} \), the sample mean of the auxiliary variable \( x \), for a given sample size (n). Also, \( n \bar{x}_{1n} \) is the total of \( x \)s in the sample. It should be clear, therefore, that minimum variance will be achieved if a sample consists of the \( n \) largest items based on the auxiliary variable \( x \) from the population, since, at the same time, the numerator in (11) would be decreased and the denominator increased. Such a sample is the optimal sample for the ratio estimator under Model (3).

The practical implication of the foregoing is that if the auditor is confident about the ratio model, he/she only needs to examine the accounts with the largest book values for a given sample size \( n \). This suggestion is clearly different from the usual practice of random sample selection and confirms our earlier assertion that a model-based approach results in a different sample design than the more familiar design-based approach. In the design-based approach, the choice of a ratio estimator was conditioned on the sample selection process while the suggested sample selection design already given is conditioned on the specified model. The optimality of the strategy is intuitively clear: if the model is appropriate in that it specifies a straight line through the origin, then one need only fix the one end point, which can be done most precisely by concentrating the sample information at that extreme point. Thus, the auditor only needs to identify the model, and optimality can be accomplished by selecting the \( n \) largest population items in the sample.

In Model (3), we assume that the variance of \( Y \) is proportional to \( x \). But, even if this condition were not to hold, the ratio estimator still would be unbiased. The following theorem proves this point.

Theorem 2. The ratio estimator, \( \hat{\pi}(0,1,x) \), is unbiased under the model of \( \xi(0,1,g(x)) \) for any variance function of \( g(x) \). A proof is provided in Appendix 2. This theorem lessens the problem for the
cases in which the variance of the audit values is not necessarily proportional to the book value.

The optimality of the sample design selecting the largest population items is conditioned on the correctness of the superpopulation model specified. Suppose that the correct superpopulation model is not the ratio Model in (3), but a regression model with nonzero intercept; will the sample design still yield an efficient, unbiased estimator under Model (3)? We now consider the bias and mean square error (MSE) of the ratio estimator when the assumed model is of the latter form. Proofs of the following theorems are contained in Appendix 2.

Theorem 3. The bias of the ratio estimator, \( \hat{T}(0,1:x) \), under the model \( \xi[1,1:x] \) is
\[
E[\hat{T}(0,1:x) - T] = \beta_o \left[ n \sum_{i} x_i / \sum_{i} x_i - (N-n) \right]
= \beta_o N(\bar{x} - \bar{\bar{x}})/\bar{x}.
\]
Its mean square error is
\[
E[(\hat{T}(0,1:x) - T)]^2 = \sigma^2(\xi(1-f)/(\bar{x} - \bar{\bar{x}}) + [\beta_o N(\bar{x} - \bar{\bar{x}})/\bar{x}]^2.
\]
(13)

One then can compare the mean square errors of the BLU estimator and that of the ratio estimator under the model \( \xi[1,1:x] \) which is a linear regression with an intercept and an error distribution as the Model in (3). The following theorem provides the BLU estimator and its MSE for this regression model with a non-zero intercept.

Theorem 4. Under the model \( \xi[1,1:x] \), the BLU estimator is
\[
\hat{T}(1,1:x) = \sum_{i} y_i + (N-n)\hat{\beta}_o + \hat{\beta}_1 \sum_{i} x_i,
\]
where
\[
\hat{\beta}_o = (\sum_{i} (y_i/x_i) \sum_{i} x_i - n \sum_{i} y_i) / D,
\]
\[
\hat{\beta}_1 = (\sum_{i} y_i \sum_{i} (1/x_i) - n \sum_{i} (y_i/x_i)) / D,
\]
with \( D = n^{\eta}\bar{x}\bar{x}_{(-1)}^{(-1)} - 1 \)
and \( \bar{x}_{(-1)} = (1/n) \sum_{i} (1/x_i) \).

Its MSE is:
\[
E[(\hat{T}(1,1:x) - T)]^2 = \sigma^2(\xi(1-f)/(\bar{x} - \bar{\bar{x}})) + \sigma^2[\xi^{2}(\bar{x} - \bar{\bar{x}})^2]/\{\bar{x}_{(-1)} n\bar{x}_{(-1)}^{(-1)} - (1/\bar{x})]\}.
\]
(14)

Comparing (13) and (14) indicates that if \( x = x \), the MSE of the biased ratio estimator \( \hat{T}(0,1:x) \) will equal that of the BLU estimator under the model \( \xi[1,1:x] \). A sample with its \( \bar{x} = \bar{x} \) can be chosen and the MSE will be that of the BLU estimator under model \( \xi[1,1:x] \).
The foregoing point deserves further emphasis. It indicates that, if the superpopulation Model in (3) (the ratio model) is not the correct model, the ratio estimator \( \hat{T}(0,1:x) \) with a sample selection design based on the largest population elements will yield a biased estimator. However, if the correct model is the regression model \( g(1,1:x) \), the ratio estimator \( \hat{T}(0,1:x) \) will yield the same unbiased estimate as the regression model, if, instead of selecting the largest population elements, the sample selection design is revised to select a sample whose mean, \( \bar{x}_s \), equals the population mean, \( \bar{x} \). This result suggests that auditors might be able to rely on a ratio estimator even if the somewhat more complex regression model is the correct one, if they “balance the sample design” by selecting a sample where \( \bar{x}_s = \bar{x} \). Note that this does not imply a random sample from the population. Of course, if the ratio model is not correct, the more general regression model with a non-zero intercept may not be correct either. As noted later, a somewhat more elaborate balancing plan may be required in that case to sustain the ratio models.

Furthermore, the bias of the ratio estimator \( \hat{T}(0,1:x) \) under a more general model \( g(h_0, \delta_1, \ldots, \delta_j; g(x)) \) is

\[
E[\hat{T}(0,1:x) - T] = \sum_{j=0}^{J} \delta_j \beta_j N \bar{x} (\bar{x}^{(j)}/\bar{x}) - (\bar{x}^{(j)}/\bar{x}),
\]

where \( \bar{x}^{(j)} \) indicates the mean \( j \)th power of \( x \), and \( \bar{x}^{(j)} \) indicates the mean \( j \)th power of the sample values of \( x \). This bias can be eliminated if a sample is chosen such that \( \bar{x}^{(j)}/\bar{x} = x^{(j)}/\bar{x} \), for \( j = 0, 1, \ldots, J \). This sample would be a balanced sample for a ratio estimator when an even more general model like a linear regression with a non-zero intercept is, in fact, the correct superpopulation model. A balanced sample would thus provide some protection from misspecification of the model.

**IMPLICATIONS FOR RESEARCH**

Auditing populations have characteristics different from those of populations with which design-based sampling strategies can most effectively deal. Large-sample theory predicts, and empirical studies support, that design-based sampling will not perform well with the sample sizes auditors are limited to for economic reasons. We have examined an approach suggested by model-based sampling and a specific ratio model. This approach of formalizing the prior information in a superpopulation model, specifying explicit criteria for evaluating estimators and then developing appropriate sampling strat-
egies appears to have the potential to deal effectively with accounting populations. It is possible that the characteristics of rare populations which lead to poor performance of conventional sampling strategies could be used as a source of information to construct superpopulation models and derive suitable sample design procedures and estimators. Currently, this prior accounting population information is used in connection with various estimators in DUS sampling, but its implications for the sample-design process have not been considered. In those cases in which the actual population does conform to the superpopulation model, sampling efficiencies would be realized. Robust sampling designs ("balanced sample designs") can be used to provide protection from misspecification of the model.

The research issue is how to utilize the substantial prior information the auditor has concerning the likelihood of accounting errors in a robust model-based design. From the previous discussions, the incorporation of explicit error distributions in sampling strategies appears to be a viable approach to overcome the rare-error problem in accounting populations. We shall continue to examine the impact of such approach on the theories and performance of model-based sampling in auditing. This paper has attempted to explain the theoretical basis of the audit-sampling problem, to review approaches to its solution and to introduce model-based sampling as a more general approach. The research issue has been stated. Additional research on model-based sampling is needed, however, to determine its ultimate benefits in auditing.

APPENDIX 1. DISCUSSION OF A MODEL APPROACH IN ANALYTICAL REVIEW

Numerous models have been proposed as aids in analytical review, for example, AICPA, SAS No. 39 (1981); Stringer (1975); Deakin and Granof (1974); and Kinney (1978 and 1979a). The explicit application of formal models in an audit may be called the model approach in auditing. A good example of such an approach is found in Statement on Auditing Standards (SAS) No. 39 (AICPA 1981), which provides a general model formalizing the relationships among the risks associated with the auditor's evaluation of internal controls, substantive tests of details, analytical review procedures, and other relevant substantive tests:

\[ UR = IC \times AR \times TD, \]

where
\[ UR = \text{the allowable ultimate risk that material monetary errors might exist in account balances or class of transactions;} \]
\[ IC = \text{the auditor's assessment of the risk that the internal accounting control system may fail to detect the existence of material errors;} \]
\[ AR = \text{the auditor's assessment that analytical review procedures or other relevant substantive tests could fail to detect the existence of material errors, and} \]
TD = The allowable risk of incorrect acceptance for substantive tests of details given that material errors are not detected by the internal control system or analytical review procedures and other relevant substantive tests.

Taking into consideration the various precautions discussed by Kinney (1988), this model provides a general framework around which the auditor can plan for the appropriate risk levels of various auditing procedures to achieve the desired ultimate risk.

Regression type models provide another example of a model approach in auditing. The Deloitte, Haskins & Sells STAR program employs such a model to assist in the analysis of financial statement accounts. STAR would support the construction of a regression model based on a firm’s past data and provide predicted values for the current period dependent variables. Stringer (1975) concluded that such a model approach could provide further insight into clients’ business operations, increase objectivity in analytical review, and result in a reduction of tests of details.

Deakin and Grauof (1974) proposed the use of a regression model to predict account balances and thus, determine the existence of material errors. Kinney (1978) also constructed regression models to predict the revenues of railroad firms. Kinney found the regression model to be reasonably accurate and economically beneficial. In another study, Kinney (1979a) proposed the integration of regression analysis with partitioned dollar-unit sampling to control risks of overstatement errors in income statement accounts. Kinney and Salamon (1982) provided empirical comparisons for the models used in STAR and those proposed by Kinney (1979a).

Underlying these regression-type models is a degree of understanding about the relationship between the dependent variables and independent variables made explicit by the model. For example, if we treat the true (audit) values as the output from a generating process of book values plus some random errors, the functional form of this generating process can be written as

\[ y = a + bx + e, \]

where \( y \) is the audit value, \( x \) the book value, \( e \) the error term, and \( a, b \) are the constant coefficients. As conjectured by Kinney (1979h), living systems have the tendency to maintain their internal equilibrium by regulating the processes in their bodies. Business organizations also have the tendency to adjust their resources toward an equilibrium. If the expected audit values can be considered to be a measure of their equilibrium, the departure of this expected audit values from book values provides some indication of possible accounting errors or irregularities. Recognition and formalization of such functional relationships between the true (audit) values and other variables (such as book values), and then utilizing this knowledge proves to be useful in analytical review procedures.

However, the model approach has not been carried over to substantive tests of detail. When statistical sampling is used in substantive tests of details, auditors typically adopt available sampling strategies from conventional survey-sampling theories. Thus, the auditor usually selects a random sample and estimates the audit value from this sample without resorting to the
functional models for the audit values and other variables. Within the conventional sampling framework, inference about audit values are made, based on the probability structure induced in the randomized sample design process. Unfortunately, due to the unusual characteristics of most accounting populations, conventional sampling strategies have failed to produce satisfactory results. Accounting researchers have tried to improve the performance of the sampling strategies with proposals of various modifications within the framework of the conventional sampling approach including the techniques of stratification and estimators using auxiliary variables. An overview of the existing approach to the integration of analytical review and conventional design-based sampling strategies in substantive tests is provided in Chart 3.

Current developments in finite population sampling propose a more explicit utilization of the functional relationship between primary variables and auxiliary variables than those considered in conventional survey sampling methods. These functional relationships are captured in a superpopulation model on which a new set of sampling strategies are derived. The use of a superpopulation model in this new "model-based sampling" appears to be in line with the growing application of the model approach in analytical review procedures. The relationship between analytical review and model-based sampling in substantive testing is summarized in Chart 4.

APPENDIX 2. PROOF OF THEOREMS

Theorem 2. The ratio estimator, \( \hat{T}(0.1; x) \), is unbiased under the model of \( \xi[0,1; g(x)] \) for any variance function of \( g(x) \).

Proof:

(A.1) \[ \hat{T} = \hat{T}(0.1; x) = (\bar{y} - \bar{x})/\bar{N_x}. \]

For Model \( \xi[0,1; g(x)] \),

(A.2) \[ Y_i = \beta x_i + \epsilon_i, \]

Let \( d_i = g(x_i) \)

\[ \bar{d} = \sum_{i=1}^{N} d_i / n; \]

\[ \hat{\lambda} = \sum_{i=1}^{N} d_i / N. \]

Then, \( E(d_i) = E(g(x_i)) \)

\[ = (g(x_i))^2 E(\epsilon) = 0; \]

\[ V(d_i) = g(x_i) V(\epsilon) = g(x_i) \epsilon^2; \]

\[ E(\bar{d}) = \sum_{i=1}^{N} E(d_i) / n = 0; \]

\[ E(\hat{\lambda}) = \sum_{i=1}^{N} E(d_i) / N = 0. \]

From (A.1) and (A.2),

\[ \hat{T} = N \bar{x}(\bar{y} - \bar{x}) / \bar{N_x}; \]

\[ = N(\bar{y} + \hat{\lambda} - \bar{x}). \]

Analytical Review

Prior period aggregate data → Construct Aggregate regression models → Predict Audit period aggregate account balance → Substantive tests of details

Audit period independent variable data

Substantive Tests of Details (Design-based Sampling Strategies)

Conventional

Audit period random sample data → Rely on Estimators

Large sample normal approximation → Estimate Audit period aggregate account balance

Probability assumptions of error frequencies & taintings → Estimate Audit period upper error bound

DUS

Analytical Review

Prior period aggregate data → Construct → Aggregate regression models → Predict → Audit period aggregate account balance → Substantive tests of details

Audit period independent variable data

Substantive Tests of Details
(Model-based Sampling Strategies)

Prior information → Construct → Regression (or other) models → Predict → Audit period aggregate account balance

Audit period nonrandom sample data

Estimators
As \( T = N\gamma = N(\beta \bar{x} + \bar{d}), \)
\[
\hat{T} - T = N(\bar{d} - \bar{d});
\]
\[
E(\hat{T} - T) = E[N(\bar{d} - \bar{d})]
= N(\bar{E}(\bar{d}) - E(\bar{d}))
= 0.
\]

So, \( \hat{T}(0,1; x) \) is unbiased under the model \( \xi[0,1; g(x)] \) for any variance function \( g(x) \).

Theorem 3. The bias of the ratio estimator, \( \hat{T}(0,1; x) \), under the model \( \xi[1,1; x] \) is
\[
(A.3) \quad E[\hat{T}(0,1; x) - T] = \beta_n \left( \sum \frac{x_i}{\bar{x}} - (N-n) \right)
= \beta_n N(\bar{x} - \bar{d}) / \bar{x}.
\]

Its mean square error is
\[
(A.4) \quad E[\hat{T}(0,1; x) - T]^2 = \sigma^2 (N(1-f)/f)(\bar{d} / \bar{x})^2 + [\beta_n N(\bar{x} - \bar{d})]^2
\]

Proof: (a) The bias for model \( \xi[1,1; x] \),
\[
(A.5) \quad Y_i = \beta_0 + \beta_1 x_i + \epsilon(g(x)) i
\]

Let \( \epsilon(g(x)) \) be \( d \) and the notations of \( \hat{d} \) and \( \bar{d} \) are the same as used in Theorem 2.

When Model (A.5) is the true model,
\[
\hat{T} = N\bar{x}(\bar{x}/x_i) = N\bar{x}(\beta_0 + \beta_1 x_i + \bar{d}) / \bar{x};
\]
\[
= N[\beta_0 + \beta_1 x_i + \bar{d}] / \bar{x}
= N[\beta_0 + \beta_1 x_i + \bar{d}]
\]
\[
T = N\bar{x}
\]
\[
\hat{T} - T = N[\beta_0 + \beta_1 x_i + \bar{d}] - N[\beta_0 + \beta_1 x_i + \bar{d}] / \bar{x}
= \beta_n N(\bar{x} - \bar{d}) / \bar{x}
\]

Then,
\[
E[\hat{T} - T] = N[\beta_0 + \beta_1 x_i - 1] + E(\bar{d}) x / \bar{x} - E(\bar{d})]
= N[\beta_0 N(\bar{x} - \bar{d}) / \bar{x} - N]
= \beta_n N(\bar{x} - \bar{d}) / \bar{x} - (N-n)
= \beta_n N(\bar{x} - \bar{d}) / \bar{x} - (N-n)
\]

(b) Mean square error
\[
E(\hat{T} - T)^2 = [E(\hat{T} - T)]^2 + V(\hat{T} - T)
= [N\beta_0 N(\bar{x} - \bar{d}) / \bar{x} - N] + V[N(\bar{d} x / \bar{x} - \bar{d})]
= \beta_n N(\bar{x} - \bar{d}) / \bar{x} + V[N(\bar{d} x / \bar{x} - \bar{d})].
\]

Let \( f = n / N \)
\[
V[N(\bar{d} x / \bar{x} - \bar{d})]
= V[\{\bar{x}/(1/f - 1)\} \sum_i d_i - (\sum_i d_i + \sum_i d_i)]
= \{\bar{x}/(1/f - 1)\} \sum V(d_i) + \sum V(d_i)
\]
\begin{align*}
= \left(\frac{\bar{x}}{\bar{y}}\right)/(1/f) - \frac{1}{f^2} \sum_i \sigma^2 x_i + \sum_i \sigma^2 x_i \\
= \left(\frac{\bar{x}}{\bar{y}}\right)/(1/f) \sum_i \sigma^2 n \bar{x}_i + \sigma^2 (N-n) \bar{x}_i \\
= \left(\frac{\bar{x}}{\bar{y}}\right)/(1/f) \sum_i \sigma^2 n \bar{x}_i + \sigma^2 (N-n) \bar{x}_i \\
= \left(\frac{\bar{x}}{\bar{y}}\right)/(1/f) \sum_i \sigma^2 N \bar{x}_i + \sigma^2 N (1-f) \bar{x}_i \\
= \sigma^2 N (1-f) \bar{x}_i ((1-f)/(1-f) \bar{x}_i) + 1 \\
= \sigma^2 N (1-f)/(1-f) \bar{x}_i.
\end{align*}

(Note: \( \bar{x} = \bar{y}, \bar{y} + (1-f) \bar{x} \))

Hence,

\begin{equation}
E(\hat{T}(0,1;x)-T) = \sigma^2 (N(1-f)/(1-f)(\bar{x}/\bar{y})) + |\beta_0 N(\bar{x} - \bar{y})/\bar{x}|^2.
\end{equation}

Theorem 4. Under the model \( \xi[1,1;x] \), the BLU estimator is

\[ \hat{T}(1,1;x) = \sum_i y_i + (N-n) \tilde{\beta}_0 + \tilde{\beta}_1 \sum_i x_i, \]

where

\[ \tilde{\beta}_0 = (\sum_i x_i/y_i) \sum_i x_i - n \sum_i y_i / D; \]
\[ \tilde{\beta}_1 = (\sum_i x_i \sum_i 1/y_i - n \sum_i (x_i/y_i)) / D. \]

Its MSE is

\begin{equation}
E(\hat{T}(1,1;x)-T)^2 = \sigma^2 (N(1-f)/(1-f)(\bar{x}/\bar{y})) + \sigma^2 [N^2(\bar{x} - \bar{y})/(\bar{x}^2 n[\bar{x} - \bar{y}] - (1/\bar{x})]].
\end{equation}

Proof: (a) The BLU estimator \( \hat{T}(1,1;x) \).

Let

\[ X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad W = \begin{bmatrix} x_1 & 0 & \cdots & 0 \\ 0 & x_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x_n \end{bmatrix}, \]

\[ \tilde{T} = \sum_i y_i + \sum_i \tilde{y}_i. \]

According to Gauss-Markov Theorem, the BLU estimator for \( \sum y_i \) is

\[ \tilde{\beta}_0 (N-n) + \tilde{\beta}_1 (N-n) \bar{x}_i, \]

where

\[ \begin{bmatrix} \tilde{\beta}_0 \\ \tilde{\beta}_1 \end{bmatrix} = (X' W^{-1} W^{-1} Y)' X^{-1} W^{-1} Y. \]

\[ = \begin{bmatrix} \sum_{i=x_1}^{i=N} n \\ n \sum_i x_i \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{i=N} y_i \\ \sum_{i=1}^{i=N} \bar{x}_i \end{bmatrix} \]

\[ = \frac{1}{D} \begin{bmatrix} \sum_i x_i & -n \\ -n & \sum_i x_i \end{bmatrix} \begin{bmatrix} \sum_i y_i \\ \sum_i \bar{x}_i \end{bmatrix}. \]
where
\[ D = n^2(\bar{x}\bar{x}^{(-1)} - 1) = \sum_i (1/x_i) \sum_j x_j - n^2. \]

Therefore,
\[ \hat{\beta}_0 = (\sum_i (y_i/x_i) \sum_j x_j - n \sum y_i)/D; \]
\[ \hat{\beta}_1 = (\sum_i y_i \sum_j (1/x_j) - n \sum (y_i/x_i))/D. \]

(b) MSE under model \([1,1;x]\).

Let \( \hat{T} = \sum_i \hat{\gamma}_i - \sum_j \hat{\mu}_i \)
\[ \hat{T} - \bar{T} = \sum_i \hat{\gamma}_i + \hat{\beta}_0(N-n) + \hat{\beta}_1(N-n)\bar{x}_i - \sum_i \gamma_i - \sum_j \mu_i \]
\[ = \hat{\beta}_0(N-n) + \hat{\beta}_1(N-n)\bar{x}_i - (N-n)(\beta_0 + \beta_1\bar{x}) + \hat{\beta}_0 + \hat{\beta}_1\bar{x}_i - \hat{\beta}_0 \]
\[ = (N-n)(\hat{\beta}_0 - \beta_0) + (\hat{\beta}_1 - \beta_1)(\bar{x}_i - \bar{x}). \]

So,
\[ V(\hat{T} - \bar{T}) = (N-n)^2[V((\hat{\beta}_0 - \beta_0) + (\hat{\beta}_1 - \beta_1)\bar{x}_i) + V(\bar{T})]. \]

and
\[ V((\hat{\beta}_0 - \beta_0) + (\hat{\beta}_1 - \beta_1)\bar{x}_i) = V(1 \bar{x}_i) V\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} \]
\[ = (1 \bar{x}_i) V\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} \]
\[ = (1 \bar{x}_i) V\begin{pmatrix} \bar{1} \\ \bar{x}_i \end{pmatrix}. \]

(\text{A.7})

According to the weighted least square theory,
\[ V\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = V((X'W^{-1}X)^{-1}X'W^{-1}Y) \]
\[ = \sigma^2(X'W^{-1}X)^{-1} \]
\[ = \sigma^2 \begin{bmatrix} \bar{x}^{-1} & n \\ n & n\bar{x}_i \end{bmatrix}^{-1}. \]

where \( \bar{x}^{-1} = (1/n) \sum_i (1/x_i). \)

Therefore, (\text{A.7}) becomes
\[ V((\hat{\beta}_0 - \beta_0) + (\hat{\beta}_1 - \beta_1)\bar{x}_i) \]
\[ = \sigma^2[1 \bar{x}_i] \begin{bmatrix} \bar{x}^{-1} & n \\ n & n\bar{x}_i \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ \bar{x}_i \end{bmatrix}. \]

(\text{A.8})

On the other hand,
\[ V(\hat{d}) = V\left(\frac{1}{N-N-n}\sum_{i}d_i\right) \]
\[ = \left(\frac{1}{N-N-n}\right)^2 \sum_{i}V(d_i) \]
\[ = \left(\frac{1}{N-N-n}\right)^2 \sum_{i}\sigma^2 \hat{x}_i \]
\[ = \left(\frac{1}{N-N-n}\right)\sigma^2 \hat{x}_j \]
\[ (A.9) \]

From (A.8) and (A.9), the MSE for \( \hat{T}(1,1:x) \) is

\[ \text{MSE}[\hat{T}(1,1):T] = \sigma^2\left(\frac{N}{N-n}\right)(1-f)(\hat{x}_j/\hat{x}_j) + \]
\[ \sigma^2\left(\frac{N^2(\hat{x}-\hat{x})^2}{N^2n[\hat{x}^{-1}-(1/\hat{x})]}\right). \]
\[ (A.6) \]

NOTES

1. Readers who are familiar with design-based statistical auditing sampling may wish to scan the first two sections.

2. Two points are relevant here. First, we also could use the error amounts, denoted by \( z = (z_1, z_2, \ldots, z_N) \), as the variable under consideration. If we denote the book value as \( x = (x_1, x_2, \ldots, x_N) \), then \( z_i = x_i - y_i \) for \( i = 1, 2, \ldots, N \), where \( y_i \) is the audit value.

Second, any specific outcome of \( Y \) is a point in this \( N \) dimensional space. There is an infinite number of points in this space and the term “super-population model” is used to describe the distribution of these points in the \( N \) dimensional space.

3. Do not confuse the accounting error as denoted by \( z \) and the statistical error term \( e \) in the model.

4. Balanced sampling considers the possibility that the super-population model could be a polynomial regression with a dependent primary variable (for example, the audit value) and an independent auxiliary variable (for example, the book value), and the ratio estimator is used.

5. One such algorithm is for the maximum difference of the absolute value between the sample cumulative distribution function on auxiliary variable \( x, F(x) \), and the population cumulative distribution, \( F(x) \), to be minimized. This is called a best-fit sample (Royall and Cumberland 1981). The procedure is (1) to rank the population units according to the size of auxiliary variable \( x \) so that \( x_1 < x_2 < \ldots < x_N \); (2) to compute \( j_k = (2k - 1)N/2n + 1 \) where \( n \) is the predetermined sample size, and \( k = 1, 2, \ldots, n \); and (3) to choose a sample \( s = (i_k = 1, 2, \ldots, n) \), where \( i \) is either \( f_i \) or \( f_i - 1 \) if \( j_k \) is an integer or \( \text{greatest integer less than } j_k \) otherwise.

REFERENCES


R. A. Leitch, John Neter, R. Plante, and P. Sinha, "Modified Multinomial


In their paper, Duke, Bailey, Ko, and Nachtsheim discuss two basic philosophies of the theory of inference for finite populations: design-based inference, the basis for most audit sampling methods currently in use, and model-based inference, proposed by the authors as an alternative approach to inference in auditing.

This discussion is organized roughly parallel to the paper. First, design-based inference is discussed along with some of its limitations. Second, a brief overview of model-based sampling is presented and its potential role in auditing is considered. Third, the authors' use of the model-based approach to derive implications for auditing is evaluated. A final section includes a summary of the main points of the discussion.

**DESIGN-BASED INFERENCE**

In design-based inference, the statistical problem is to design a joint-sample selection/sample-evaluation strategy that will support a conclusion about some characteristic of the population (such as total
audited value) with some specified level of statistical confidence. Randomness in this approach to inference arises from the random selection of sample units from the finite population, which is taken as fixed. For example, 100 accounts might be randomly selected to form a confidence interval for the total of the audited balance in a population of 1,000 accounts. In design-based sampling, such confidence intervals do not require any assumptions about the structure of the finite population and the inferences are, in theory, distribution-free. However, here, as in most cases in statistical inference, the price of freedom from assumptions is a potential loss in efficiency. Thus, what appears to be an advantage of the design-based approach may actually be an important drawback (Smith 1976 and Basu 1971). Simply stated, an approach to inference which is designated to work well for all audit populations may not be most efficient for any single audit problem.

In the first section of their paper, the authors review several problems that may arise when design-based sampling approaches are applied to “rare populations”; that is, populations in which errors occur only rarely. These problems arise for those design-based methods that implicitly assume that a sample statistic (such as the sample mean) is distributed approximately normally. For a sample with a small number of errors, the approximation (based on large-sample theory) may be imprecise. However, the fact that large-sample properties do not hold in small samples cannot be taken as a failure of the theory of design-based inference. Rather, this is an empirical problem which arises in applications with small sample sizes. Though this well-known empirical problem does indicate a need for different techniques (possibly, other design-based techniques) for application to rare populations, it does not imply that design-based inference must be abandoned.

In the next section, the authors note two approaches that have been used to develop design-based techniques which also can be used for rare populations. The first approach, which the authors refer to as compensating for coverage, uses the basic framework of large-sample theory, but modifies the calculation of one or more of the terms used to calculate the confidence limit. Examples of this approach are Frost and Tamura (1982) and Garstka and Ohlson (1979). The second approach, which the authors refer to as direct estimation of an upper bound, departs from the framework used for large sample theory and attempts to construct a bound based on some alternative framework. Examples include Stringer (1963), the theoretically elegant multinomial approach of Fienberg, Neter, and Leitch (1977),
and Neter, Leitch, and Fienberg (1978), and the relatively simple, but apparently effective, moment bound of Dworin and Grimlund (1984).

At most, the evidence reviewed in the first two sections of the paper indicates that some design-based approaches perform poorly in rare populations and also that there are ongoing attempts to develop new design-based approaches. Since none of the design-based approaches developed thus far solve all of the problems facing the auditor, we should be open to alternatives, including alternative approaches to the theory of inference in finite samples. However, in evaluating any alternative theory, it is important to consider specific empirical problems which might arise in auditing, especially since the empirical problems which arise for some design-based methods provided the motivation for the search for an alternative theory.

MODEL-BASED INFERENCE

In model-based inference, the finite population is assumed to be the product of a superpopulation distribution or generating model. In this approach to inference, a sample is used to obtain evidence about the parameters of the assumed generating model and estimates of the parameters form the basis for inferences about the remainder of the population. Thus, the auditor's problem is to use observations from the population to estimate the parameters of the generating model. Inferences about the remainder of the population (the unobserved portion) are based on the estimates of the population parameters. For example, a sample of 100 accounts (from a population of 1,000) might be used to estimate the parameters of the model relating audited values to book values introduced by the authors in the third section. Given the parameter estimates and the book values of the remaining 900 accounts, the auditor can draw inferences about the audited value of the entire population.

The model-based approach has considerable intuitive appeal. The assumption that there is some (unknown) model generating errors (or balances) in a client's records seems a potentially useful way to characterize the finite population. However, the critical determinant of the performance of this approach will be our ability to find models that are appropriate to the environment of auditing, and there is reason to expect that this will be a difficult task. As the authors suggest, a client's accounting system and system of internal controls serves as a filter for accounting errors. This filter is exceedingly complex and it may be unreasonable to assume that it can be
characterized by a single generating model. For instance, it may not be reasonable to assume that the error-generating model for small accounts has the same form and parameters as the error-generating model for large accounts. Similarly, the error-generating model for one class of transactions may differ from the model for other classes of transactions. On a conceptual level, model-based inference can accommodate many different generating models for a population. However, in applications, it would be impractical to have more than a very small number of models and parameters to be estimated.

The authors consider a generating model for which the audited value, $Y$, is the sum of a linear function of book value, $x$, and an error term involving $x$:

$$y_i = \beta x_i + e(x_i)^{\alpha} \text{ for } i = 1, 2, \ldots, N.$$ 

A careful evaluation of this model reveals four issues which must be addressed before model-based inference can be applied in auditing.

First, although it is analytically convenient, it is not clear why the error term should be assumed proportional to the square root of the corresponding book value. This assumption is not obviously unacceptable, but neither is it so appealing as to be acceptable without further justification. The main purpose in making this point is to emphasize the dependence of the model-based approach on the assumptions of the model, including the assumed distribution of the error terms.

Second, the assumption that the covariance of the error terms is zero is suspect, given the nonrandom nature of the error generation process in auditing. Third, this model illustrates the fact that model-based inference in auditing is dependent on distributional assumptions and/or approximations in much the same way as design-based inference. Suppose that the specified model is correct. Then the auditor's problem is to draw a sample, estimate $\beta$, and use the estimate of $\beta$ to estimate and place a bound on the audited value of the remainder of the population. Thus, the problem is not simply to estimate the audited population balance, but to calculate a confidence bound for the audited balance. However, it is not possible to construct exact confidence bounds without making an assumption about the distribution of the error terms. This assumption is necessary because the confidence bounds on the audited population balance depend on the distribution of the estimate and on the distribution of error terms. Further, even when the distribution of error terms is known, the distribution of parameter estimates may be analytically intractable.
Thus, it is not sufficient for the auditor simply to find an unbiased estimator, nor even the minimum variance unbiased estimator, of the audited population balance. Instead, the auditor must find the entire distribution of the estimator to construct bounds on the audited value of the population. Without the entire distribution, users of model-based sampling may be forced to rely on large sample approximations much like those used in the design-based approaches criticized by the authors.

Finally, given the authors' own description of the auditing environment, even the form of the model seems unreasonable. The model posits that the audited value of every account is equal to a parameter \( \beta \) times the book value of the account plus an error term with expectation zero. However, as the authors emphasized in earlier sections, many audit populations are rare populations with only a small number of errors. In rare populations, the correct model for all but a few items has \( \beta \) identically equal to 1 and an error term which is identically 0. It seems unlikely that a model which is incorrect for most of the observations to which it is applied will prove useful for inferences about the audit population.

**IMPLICATIONS FOR AUDIT SAMPLING**

In the final section, the authors present several theorems from Royall and Herson (1973) on optimal estimators and sampling strategies. The theorems derive the minimum variance linear-unbiased estimator of the audited balance for the model just described. However, as noted previously, knowledge of the optimal unbiased estimator does little to solve the auditor's problem of specifying a confidence bound for the population.

The theorems also are used to draw some conclusions about sampling strategies which, though valid in the artificial world of the assumed generating model, have limited applicability to the auditing environment. For example, the authors conclude that when all account balances are assumed to be generated by the model already discussed, the optimal sampling strategy is to select only the largest items for sampling. However, there are at least two reasons to question this conclusion in auditing. First, errors generated in an auditing context may not be independent of the account book value. Errors of different types and magnitudes are likely to occur with different probabilities for different book values, due either to characteristics of the accounting system or to human intervention. Because of these factors, few auditors would feel comfortable conducting audits in which only the largest account
balances are audited. Second, this result ignores the cost of auditing accounts; unless the cost of auditing an account is independent of the book value of the account, the conclusion is invalid.

Theorem 2 establishes that unbiasedness of the estimate is not dependent on correct specification of the variance of the error terms. However, the auditor's problem is to construct confidence bounds, not simply to identify unbiased estimators; and correct specification of the confidence bound is dependent on correct specification of the variance of the error terms. Moreover, misspecification of the covariance structure of the error terms will lead the auditor to a misspecified confidence bound, even in the convenient case where the stochastic error terms are distributed normally.

Theorems 3 and 4 present the bias and mean squared error of estimates based on misspecified models. Balanced sampling is proposed as a way of correcting for bias due to misspecification. The motivation for this proposal in the auditing context needs further consideration. It would seem preferable, for example, to deal with the problem of bias due to misspecification directly, by correctly specifying and estimating a more general model, rather than indirectly, by attempting to balance the sample.

CONCLUSION

This paper makes an important contribution by initiating a careful reexamination of the philosophical foundations of statistical inference in auditing. Model-based inference, which provides a means of implicitly incorporating prior information about the audit population, is an interesting alternative to design-based inference. However, model-based inference may be subject to many of the same empirical problems which arise in design-based sampling. Further, many of the readily available analytical results for model-based inference do not directly address the problems facing the auditor. Before model-based sampling can make a contribution in future auditing research and practice, careful consideration must be given to the assumptions of model-based inference and the suitability of the model-based framework for the auditing environment.

REFERENCES

Lowell Dworin and R. A. Grimlund, "Dollar Unit Sampling for Accounts
218–41.
S. E. Fienberg, John Neter, and R. A. Leitch, "Estimating the Total
Overstatement Error in Accounting Populations," Journal of the American
P. A. Frost and Hirokuni Tamura, "Jackknifed Ratio Estimation in Statistical
103–20.
S. J. Garstka and P. A. Ohlson, "Ratio Estimation in Accounting Populations
with Probabilities of Sample Selection Proportional to Size of Book
John Neter, R. A. Leitch, and S. E. Fienberg, "Dollar Unit Sampling:
Multinomial Bounds for Total Overstatement and Understatement
R. M. Royall and Jay Herson, "Robust Estimation in Finite Populations, I,"
pp. 880–89.
K. W. Stringer, "Practical Aspects of Statistical Sampling in Auditing,"
Proceedings of the Business and Economic Statistics Section, American Statistical
Model-based Sampling and Its Implications for Auditors: Discussion

Duke, Bailey, Ko, and Nachtsheim have established two main premises for pursuing an investigation of the application of model-based sampling to auditing. The first is that accounting populations are “rare” (that is, they contain few errors) and, thus, they may be hard for auditors to work with. (I prefer, however, to use the term low-error-rate populations to denote this characteristic of accounting populations.)

The second motivational premise is the inadequacy of presently used statistical techniques. Specifically, the authors point out that certain statistical estimators do not behave in a reliable manner when normal distribution theory is applied, with what auditors would consider reasonable sample sizes, to rare populations. Model-based sampling, the authors argue, should be adopted, or at least explored, as a “solution” to these problems.

In discussing this paper, I will focus on three issues:

1. Whether conventional statistical techniques are, in fact, unsatisfactory in assessing low-error-rate populations.
(2) Whether the notion of model-based sampling is adequately explained in the paper.

(3) One context in which I believe the pursuit of model-based sampling might be useful.

DEFENSE OF CONVENTIONAL TECHNIQUES

The Neter-Loebbecke (1975) study of statistical estimators states that, "It is clear from this study that no one statistical procedure is optimal under all circumstances. However, the study has also shown that, for all the populations and error patterns considered, there is no situation where at least one technique is not reasonably effective." Neter-Loebbecke found that with regard to low-error-rate populations, stratified mean-per-unit estimation provides both efficient and reliable results, in direct contradiction to the authors’ premise. Duke et al. did not, however, include stratified mean-per-unit estimation in their discussion.

At the time the Neter-Loebbecke study was conducted, dollar-unit sampling was not as developed as it is today. Consequently, that method was not investigated, in depth, in the Neter-Loebbecke (1975) study. However, even from its rudimentary treatment in that study, it was clear that dollar-unit sampling would be very reliable with low-error-rate populations, and as efficient as attribute sampling, on which it is based. Attribute sampling is the most widely used statistical sampling method in auditing, and is, generally, considered to be acceptably efficient.

During the last ten years, much research on and development of dollar-unit sampling has been accomplished. The reliability of the method has been sustained without question, and heuristics have been developed to provide added efficiencies. Duke et al. recognize the existence of dollar-unit sampling, but fail to give it its due. They also neglect a very important aspect of statistical methods — that statistical methods provide information that indicates when their use may not be well advised. For example, suppose an auditor takes a random sample under a plan to use difference or ratio estimation and finds very few errors. Such an outcome suggests that the auditor not proceed with either of those evaluation methods, but with some appropriate fallback procedure. Admittedly, this is not efficient; however, it is effective, and with good planning, will not occur frequently.

So it seems to me that the authors came across an idea — model-based sampling — and were intrigued by it. They then set out to
justify applying it to auditing by setting up conventional statistical sampling methodology as a straw man. Although I am critical of their motivation, I do believe audit model-based sampling is worthy of research, but for a reason different from that put forward by the authors.

MODEL-BASED SAMPLING

Godfrey and Andrews wrote a paper at the University of Michigan in the late 1970s that suggested a model-based approach. In that paper, they proposed that the auditor attempt to model the underlying population when selecting a statistical procedure for the audit of that population. The basic process encompassed by the Godfrey-Andrews approach is that the auditor would (1) look at the nature of the audit population; (2) select from a predefined set of population types; (3) apply the appropriate statistical procedure; and (4) develop a confidence interval that would be composed of the joint probability that the interval contains the true population value and the selected population type is the correct one. Such an approach incorporates the notion of feedback available with statistical methods which already was mentioned.

As I read the authors’ discussion of model-based sampling, the following process would be used: (1) The auditor would look at the population and model it. I assume that this may eventually involve selection from a series of standard models. (2) The auditor would then design a nonstatistical sample that would be optimal for the modeled population, except that it would be “balanced” to compensate for some degree of model misspecification. (3) The sample would be chosen and a conclusion would be reached. It is not clear; however, how this would be done, nor is it clear if the risk that the model is greatly misspecified would be estimated.

Although the paper does not clearly explain how the authors’ version of model-based sampling works, it seems to me that there are four important aspects that need to be controlled for the approach to be at all rigorous: (1) the extent and accuracy of the auditor’s general prior knowledge of the population being sampled; (2) the specific application of this knowledge to the circumstances; (3) the accuracy of the auditor’s “screening” knowledge; and (4) the adequacy of the auditor’s “screening” knowledge.

These are all planning steps. They must be emphasized in developing the use of model-based sampling because of the absence of a sampling feedback mechanism.
USEFULNESS OF THE MODEL-BASED APPROACH

In the third section of their paper the authors state, “Balanced sampling is neither random nor judgmental. Rather, it is a purposive sampling selection procedure with a specific optimization goal.” I would argue that this statement is a nice piece of semantic elucidation but that, in the absence of clearer and more complete information, model-based sampling is a form of judgmental sampling.

But wait! That is not meant to be a criticism, but rather a tip-off to the usefulness of research on model-based sampling. My interpretation of statistical and nonstatistical (that is, judgmental) sampling as “competing” methods is that one uses judgmental sampling when knowledge of the population is sufficient to identify the distribution of potential errors with reasonable certainty and the related sample size, in examining items likely to contain potential errors, is efficient. One uses statistical sampling when knowledge of the population is not sufficient to identify the distribution of potential errors with reasonable certainty, or when it is particularly efficient.

My concern over the years has been that auditors who use the judgmental approach do not seriously or skillfully incorporate the requisite knowledge into the sampling process. The authors are suggesting a way to do this. The control of risk is still subjective; however, the methodology may provide a discipline that could be effective in controlling risk at a lower level.

In conclusion, I would observe that the Duke et al. paper brushes across an idea — improved judgmental sampling — that I believe is not only very useful, but also very important. Thus, I would encourage the authors to continue their efforts with some appropriate redirection.

REFERENCES


Model-based Sampling and Its Implications for Auditors: A Reply

In our paper, we attempted to adapt the statistical concept of model-based sampling to the peculiarities of audit populations. Our objective was to investigate the applicability of this concept as an alternative to design-based sampling, which now dominates audit sampling. We feel that examination of alternative approaches to methods currently used by practicing auditors is healthy and in keeping with the spirit of academic research.

We thank the discussants for their thoughtful comments and we agree with most of the points they have raised. As both discussants correctly observe, our paper did not demonstrate the clear advantage of model-based sampling. However, this does not mean that advantages do not exist. We interpret their comments as encouraging us to continue this line of research in the hope of providing a better understanding of the nature of model-based statistical inference. We particularly appreciate Burgstahler’s comments and their implications for future research in this area. However, we feel compelled to reply to two specific issues raised by Loebbecke.
First, Loebbecke questions whether present statistical methods used in auditing are unsatisfactory. We have pointed out some of the problems that exist in using statistical methods in auditing as a motivation for the examination of model-based sampling. But, Loebbecke takes the position that present methods are satisfactory and hence, presumably, there is no need for further research in this area. We stand by our original position that there still exist some unsolved problems in the area.

However, considering the statistical inference problem from a model-based point of view (which provides new insight), we have made a research contribution whether or not the insight leads to improved audit efficiency. Of course, we also hope that such insight would lead to improvements in audit practice.

Second, Loebbecke refers to model-based sampling as being based on nonstatistical samples. This is an unfortunate choice of words. It implies that model-based sampling is nonstatistical. In fact, model-based sampling is well founded in statistical inference theory and pre-dates design-based sampling. It is the dominant form of statistical inference for certain applied statistical problems. Thus, one strength of model-based sampling actually is the ability to reference statistical theory in discussing its appropriateness.

Loebbecke apparently has confused the idea of a model-based sample, which is not selected at random from a finite population, with a nonstatistical sample. Thus, he has equated a model-based sample with what is currently referred to as a judgmental sample. Judgmental sampling, as the term is popularly used, is nonstatistical. That is, a valid statistical inference is not theoretically supported by such an approach to sample selection. On the other hand, model-based sampling is soundly rooted in statistical theory. A valid statistical inference can be derived from a sample selected purposefully, based on a superpopulation model. The sample selection is derived from the model, resulting in a nonrandom but objectively selected sample that is statistically valid and is not subject to an individual auditor’s biases.

Burgstahler correctly points out that model-based techniques could be subject to the same problems associated with rare populations that arise in design-based methods. We agree, but hope that a model of rare populations might be found that would allow the auditor to use this knowledge of the population in the development of a sampling strategy. We are continuing to investigate models that could be appropriate for audit populations. We thank the discussants for pointing out to us some of the pitfalls along the way.
Session
TWO
On the Fundamental Nature of Professional Opinions: The Traditional, Bayesian, and Epistemic Methods of Attestation

... auditing has borrowed, and will probably continue to borrow, from other fields what it needs to elaborate its method and fulfill its function. However, there must be careful selection of those ideas and procedures which can be of assistance. Generally speaking, successful adoption of this kind requires an understanding of the nature of the borrowed tools. Rarely are ideas and methods in their fields such that they can be accepted without some modification. (Mautz and Sharaf 1961, p. 15-16)

The contemporary auditing literature offers two major methods for performing the attest function, the traditional and the Bayesian. A major statement of the traditional method, which is founded upon Neyman-Pearson statistical methodology, is provided by Elliott and Rogers (1972), and a robust discussion of the Bayesian method is provided by Scott (1975). In this paper, we examine the philosophical foundation of each of these methods in an attempt to understand

Editors' Note: This paper was anonymously refereed prior to acceptance.
how they relate to extant views of audit attestation. A comparison of the underlying logic of the respective methods with the fundamental nature of the audit process shows that each has engrained several characteristics which may be problematic. We then introduce a third method of attestation, the epistemic method, and contrast the merits of these three methods for the performance of the audit attest function.

Our analysis proceeds as follows. The first section ends with a brief overview of each method of attestation. A detailed account and critique of the foundations of the traditional and Bayesian methods are provided in the second and third sections, respectively. The fourth section presents a class of methods which are alternatives to either the traditional or the Bayesian method of attestation. The last section provides concluding remarks.

THE AUDITOR'S PROBLEM

In performing the attest function, the auditor faces a two-part problem which may be characterized as follows. First, the auditor performs an investigation, gathers evidential matter, and then makes an inference about the fairness of the financial statements under investigation. Second, on the basis of the evidence and the related inference, the auditor selects and renders a written report. A resolution of the auditor's problem, thus, requires two processes of logic: (1) a logic of inference, which governs the formation of a belief about the fairness of the financial statement and (2) a logic of decision, which governs selection of a written opinion. At issue herein are the capacities of the traditional, Bayesian, and epistemic methods of attestation to facilitate resolution of the auditor’s problem. The formal logic underlying these alternative methods of attestation differs significantly on two points: (1) the nature of the auditor’s inference task and (2) the means of selecting the written opinion for inclusion in the audit report.

The Neyman-Pearson and Bayesian methodologies represent two fundamental views of (inductive) inference. The cognitivist view, which is associated with the traditional method, holds that inference terminates with the acceptance of particular hypotheses as true and rejection of others as false. The behavioralist view, which is associated with the Bayesian method, holds that inference leads only to weighting of hypotheses via the assignment of probabilities, and rejects, as improper, the acceptance and rejection of hypotheses. Niiniluoto and Tuomela (1973) characterize the two views as follows:

The cognitivists defend the indispensability of inductive rules which
allow for the acceptance or rejection of hypotheses, and conceive of scientific knowledge as a body of hypotheses which are rationally accepted as (presumably) true. The behavioralists hold that inductive inference leads only to an assignment of probabilities to hypotheses and that these probabilities can be utilized by the practical decision maker. (P. 148)

Both cognitivists and behavioralists regard inference as a guide to decision making. There is, however, considerable distinction between the views of cognitivists and behavioralists. Cognitivists (implicitly) suggest that decisions be made in accordance with beliefs (that is, accepted hypothesis), arrived at via inference. Behavioralists require that decisions be made in accordance with partial beliefs (that is, probabilities), but provide no means of belief formation as part of the inference process. Salient characteristics of the traditional and the (subjectivist) Bayesian methods of inference and decision are compared in the Table.4

It should be noted that proposals to replace the traditional method of performing the attest function by Bayesian methods implicitly advocate a radically different view of inference and decision, and hence, would alter the fundamental nature of the audit opinion. This point is brought to light when the auditor's attestation problems associated with each method are examined. Representing the traditional method (and its associated cognitivist view), Elliott and Rogers (1972) specify the auditor's problem as follows:

The overall objective of an auditor in carrying out the attest function is to make reasonably certain that statements examined by him are materially correct. He therefore wishes to find material errors in the financial statements if they exist. To do so he employs numerous procedures, one of the most important being the examination of documentary evidence. (P. 46, emphasis added.)

From the behavioralist camp, Scott (1973) characterizes the auditor's problem as follows:

In brief, the auditor is engaged by the firm to report whether, in his opinion, that firm's financial statements fairly present its financial position and results of operations, in conformity with generally accepted accounting principles. To do this, the auditor examines the firm's accounting system to see whether the various records are complete and accurate. (Pp. 305-6, emphasis added.)

Each characterization is stated in two parts. The first is an assertion of the auditor's purpose, and the second specifies the means used to meet that purpose. While the second part of each characterization recognizes that the auditor collects evidential matter, there is disa-
<table>
<thead>
<tr>
<th>Inference Method</th>
<th>Traditional Method</th>
<th>Bayesian Method</th>
<th>Epistemic Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>The theory of hypothesis testing (adequacy criterion)</td>
<td>Probability revision via Bayes Theorem (balancing criterion)</td>
<td>The Bernoulli Principle (optimization criterion)</td>
</tr>
<tr>
<td>Fundamental analytical device</td>
<td>The convention of classical statistics (of test selection)</td>
<td>Bayes Theorem</td>
<td>Expected Epistemic Utility Function</td>
</tr>
<tr>
<td>Cognitive artifact</td>
<td>Conclusions about a class of hypotheses and hypothesis tests</td>
<td>Posterior Probabilities</td>
<td>Conclusions about specific hypotheses</td>
</tr>
<tr>
<td>Decision Method</td>
<td>As if</td>
<td>Bernoulli Principle</td>
<td>As if</td>
</tr>
<tr>
<td>Major property</td>
<td>Informal</td>
<td>Formal</td>
<td>Informal</td>
</tr>
<tr>
<td>Cognitive artifact</td>
<td>An opinion</td>
<td>A report-type</td>
<td>An opinion</td>
</tr>
</tbody>
</table>
greement about the auditor's objective in each first part, reflecting the different views of inference and action-choice held by cognitivists and behavioralists, respectively. Elliott and Rogers specify the auditor's goal as achieving reasonable certitude that the statements under examination are correct. That is, the auditor is to judge whether the financial statements are materially accurate. Contrariwise, Scott specifies the auditor's goal as the selection of a report. For Elliott and Rogers there is a logical connection between evidence and belief, whereas for Scott there is a logical connection between evidence and action-choice (that is, report selection).

In the remainder of the paper we examine the foundations of the traditional and Bayesian methods of attestation, and show that neither provides a complete method for resolving the auditor's problem. In so doing we detail flaws in the foundations of the traditional and Bayesian methods with respect to this problem. Finally, we propose a class of methods which overcome these flaws.

THE TRADITIONAL METHOD OF ATTESTATION

The traditional method of attestation consists of two processes, an inference process, which yields judgments, and a decision process, which yields an audit report. The inference process is based upon the Neyman-Pearson theory of statistical hypothesis testing. The decision process is simple, direct, and informal — the auditor simply reports the judgment reached via the inference process.

Hypothesis Testing under the Traditional Method

Application of the traditional method of attestation requires specification of two hypotheses, selection of a significance level, and application of a test. A proper hypothesis specification is

\[ H_0: \text{The financial statement amount is not materially in error}; \text{ and} \]
\[ H_1: \text{The financial statement amount is materially in error}. \]

The significance level typically is set at 1 or 5 percent (see Elliott and Rogers 1972, pp. 48-49). Then, depending upon the specific properties of an account balance and the related control system, the auditor selects and performs an audit test. The traditional method of attestation presupposes that the results of such a test resolve the auditor's judgment problem vis-à-vis the account balance in question. Such a presupposition is invalid, however, for two reasons. First, the significance level of a test is typically interpreted to be a probabilistic measure of the auditor's degree of certitude in the accepted hypothesis, although such an interpretation is problematic upon
completion of a test. Second, the results of the test conducted by the auditor actually are not applicable to the specific financial statement account balance in question.

The Neyman-Pearson theory of hypothesis testing recognizes two types of error. Error of the first kind consists of rejecting a true null hypothesis, whereas error of the second kind consists of accepting a false null hypothesis. The probability of committing an error of the first kind is called the significance level of a test. The capacity of a test to discriminate between the null and alternative hypotheses is called the power of the test. The Neyman-Pearson theory of hypothesis testing is governed by the convention of classical statistics: given a specified significance level, the best test is the test with maximum power. This convention also can be stated as follows: one selects the maximum risk one is willing to take for making an error of the first kind, and then selects the test that minimizes the probability of committing an error of the second kind.

Significance level, however, "cannot be construed as the probability of error in our results; in the case at hand our results either lead us into error or they do not. There is no question of probability" (Kyburg 1974, p. 32). Further, a significance level is not, and cannot be interpreted as, a measure of the degree of credibility of a hypothesis on the basis of evidence. (For auditors, significance level is a measure of how likely a fair balance is to be rejected, prior to gathering evidence, and given that the auditor has minimized the probability of accepting a materially inadequate balance). Indeed, the Neyman-Pearson theory of hypothesis testing is predicated on the view that it is impossible to develop a (scientifically) defensible theory of testing which incorporates a notion of the credibility of a hypothesis on the basis of evidence.

Recall, however, that the Neyman-Pearson theory provides only power and significance level as measures which are interpretable as probabilistic measures of the degree of certitude that an account balance is not materially in error. Further, recall that power is a measure of the capacity of a test to discriminate between the two types of error, not a measure of the strength of connection between the evidence and hypotheses. Thus, the Neyman-Pearson theory is, in principle, incompatible with cognitivist logic. Moreover, significance level must be misinterpreted if the auditor’s problem is to be resolved by the traditional method of attestation.

We now turn to a related issue. Significance level and power are used in the Neyman-Pearson theory to select a test for a pair of hypotheses of a particular kind. The results of a test, therefore, are
applicable only to that kind (or class) of hypotheses, and are of no immediate relevance to the particular hypotheses at hand (that is, "the account balance under investigation is (is not) fair"). The traditional method employs probabilities as a guide to belief formation. However, "probabilities, construed as measures in (the Neyman-Pearson) view, can only be calculated relative to the relevant hypothesis; and even the probabilities thus calculated do not characterize the particular inference or decision made, but only that kind of inference on that kind of decision" (Kyburg 1974, p. 33). Computations of probabilities of error (of either the first or second kind) are, thus, computations of probabilities of error, given that the same kind of test is repeated: they are probabilities of long-run error.9

The traditional method, as noted earlier, requires that the auditor be reasonably certain that the particular financial statement under investigation is not materially in error. But the auditor does not engage in long-run repetition of the same kind of test. The concept of reasonable certitude implicit in the traditional method, therefore, is incompatible with the long-run probabilities of error inherent in the Neyman-Pearson theory. Consequently, the Neyman-Pearson theory of inference is inadequate as a (cognitivist) method of performing the judgment component of the attest function as specified by traditional method.10

Clearly, the individual auditor bears the direct risk of an erroneous audit. The traditional method, via its Neyman-Pearson theory of inference, is vexing in this regard because it is based on the view that the risk relevant to statistical testing is the long-run risk of admitting erroneous information to the body of public knowledge. The traditional method is based upon a theory of inference which can be "seen to be a self-correcting method, that, if indefinitely followed, must in the long-run lead the (scientific) community, although not the individual reasoner, indefinitely closer to the truth" (Edwards 1967, p. 177). The auditor, as the individual reasoner, is not well served by an inference method that passes on risk to the community when, in fact, he/she actually is made to bear that risk. Nor can he/she take comfort from a method of risk assessment that measures, at best, how well he/she would do, in the long run, in replicating a particular test or class of audit tests, because the auditor performs such tests only once as a basis for forming an opinion.

This problem can be alleviated, however, by employing an inference method that overtly accounts for the auditor's bearing of situation-specific risk. The Bayesian method is not problematic in this regard and, therefore, if not for other difficulties (discussed below), would
seem to be a natural replacement for the clearly flawed traditional method.

**Action-choice Selection under the Traditional Method**

If, via judgment, the auditor formulates a belief, then acting as if the belief were true would be optimal. In the traditional method of attestation the connection between evidence and belief is formal and structured, while the connection between belief and action is informal and unstructured. In short, the traditional method employs no formal logic for action-choice.

In the auditing context the decision process yields a report. Using the traditional method of attestation, the (written) report publicly reveals the auditor's judgment concerning the material accuracy of the financial statement as determined by the inference process. A user of an audit report can only presume that the auditor acted honestly, and revealed what he/she actually believed to be the truth. Such a presumption is warranted by the auditor's adherence to ethical rules which demand integrity (Code of Professional Ethics, 1975, p. 18).

**Using the Traditional Method**

The traditional method of attestation ends with issuance of a report which reveals the auditor's actual beliefs about the financial statements under investigation. However, since the inference process of the traditional method of attestation is based on classical statistical conventions, it cannot yield the auditor's actual belief about the particular financial statements at issue. The question remains as to what the auditor really accomplishes by applying the convention of classical statistics. Any application of the Neyman-Pearson theory must be consistent with the foundation of that theory and should conform to the empirical (frequentist) view of probability. To do so, however, the auditor would have to specify a long-run sequence of events, the relative frequency of which (in the limit) is a probability (Kyburg 1974, p. 175).

A candidate sequence in the audit context is the sequence of all possible random samples (of a given size) that could be selected and observed by the auditor prior to accepting or rejecting a statistical hypothesis. The result of a single test based on one such sample is acceptance or rejection of a hypothesis about the financial statement balance in question. As noted previously, however, the Neyman-Pearson theory does not associate probabilities of error with such decisions. Thus, it provides no measure of degree of certitude in
association with such an acceptance or rejection decision. The convention of classical statistics instead would lead the auditor to accept (or reject) the statistical hypothesis at hand. Having done so, the auditor possesses no measure of certitude concerning the probability that his decision is correct. That is, adherence to the convention of classical statistics generates a report about the presence or absence of material error, but not a likelihood that error is present.

The auditor may choose to ignore this point and set about applying the classical theory of statistical inference as if it were compatible with the traditional method of attestation. However, any attempt to apply significance level as a measure of certitude associated with his/her decision about the presence or absence of material error will remain inappropriate for other reasons. Strictly speaking, attestation does not involve tests of simple hypotheses. Rather, the auditor faces composite null hypotheses which must be tested against composite alternative hypotheses. For example, the primary issue in substantive testing is whether the amount of error is material or immaterial.

The auditor can of course reformulate the problem as a test involving one simple hypothesis and a composite alternative. For example, the auditor could test the hypothesis that the financial statement is in error by precisely the material amount against the composite alternative hypothesis that the financial statement is in error by less than the material amount. On the other hand, the auditor might state the problem as a test of the composite null hypothesis, that the statement is in error by a material amount (Roberts 1974, p. 52). In either reformulation the auditor faces a test involving at least one composite hypothesis. Composite hypotheses are problematic, however, since an exact calculation of both significance level and power can be accomplished only for tests involving simple null hypotheses against simple alternative hypotheses (Kyburg 1974, p. 33). No measure exists, therefore, in the attestation context (even after reformulation) which could be used as a measure of degree of belief or certitude.

THE BAYESIAN METHOD OF ATTESTATION

The Bayesian method of attestation is a direct application of the subjectivist view of inference and decision to auditing. It consists of two parts: (1) an inference process based upon Bayes' theorem as a means to revise subjective probabilities and (2) a decision process based upon the Bernoulli principle of minimizing expected loss. Bayesian inference, consistent with the behavioralist view, ends in the
assignment of probabilities to hypotheses. The decision process utilizes these probability assignments, as discussed below, to resolve the auditor's practical decision problem of selecting an audit report.

**Using the Bayesian Method**

Subjectivist Bayesian inference is properly concerned with the use of evidence to revise personal probability assignments. Probability revision is governed by Bayes' theorem which states that if an individual initially holds the probability \( P(H) \) in hypothesis \( H \), then given evidence \( E \), he/she will revise his probability to \( P(H|E) \), where \( P(H|E) = \frac{P(H)P(E|H)}{P(E)} \). The determination of \( P(H) \), \( P(E) \), and \( P(E|H) \) are, however, problematic in virtually all substantive audit tests.

A subjective probability, such as \( P(H) \), is a measure of the (personal) degree of belief an individual has about the occurrence of a realizable event.\(^{12}\) Determination or construction of \( P(H) \) is a nontrivial practical problem with respect to coherence (that is, conformity with the probability calculus).\(^{13}\) Further, one of the major presuppositions of the subjectivist interpretation of probability is that the individual assessing prior probabilities must be willing "to accept bets of any size, on either side with respect to any combination of events" (Kyburg 1974, p. 95) for which probabilities are being assigned. But, such an assumption in the audit context seems, at best, gratuitous for several reasons. They include (1) the very real prospect of ruin faced by the auditor, and (2) the de facto absence of an adversary betting relationship without which there is no ground for requiring the individual to maintain degrees of belief that satisfy the probability calculus.\(^{14}\)

A defender of the subjectivist view might argue that if the auditor's beliefs are incoherent, then users of the financial statements will find it in their own best interest to bring suit against the auditor. The existence of this potential adversary will, thus, force the auditor to be coherent (that is, it is in the auditor's own best interest to hold coherent degrees of belief). Such an argument is not without some merit. Indeed, the auditor might estimate his degrees of belief by considering a hypothetical system of bets made against a presumed diabolical adversary. This approach is implicitly suggested by Kinney (1975, p. 120). Thus, an auditor who employs the Bayesian method holds degrees of belief which are estimates of probabilities arrived at on the basis of hypothetical bets.

Having discussed difficulties associated with \( P(H) \), we now consider problems confronted by the auditor when dealing with the remaining
elements of Bayes' theorem, that is, \( P(E) \) and \( P(E|H) \). The Bayesian auditor must assign subjective probabilities to hypotheses of the form, "the financial statement under investigation is in error by \( \$ \)—— amount." This form of the hypothesis is appealing in the audit context because it allows the states (and hence the outcome mappings) to be partitioned. Such partitioning is sensitive to different materiality estimates as well as to the presence or absence of material error. An alternative hypothesis statement of the form, "the financial statement is in error by at least a material amount," might otherwise be sufficient (and apparently is sufficient for those auditors who attempt to employ the traditional method of attestation).

The resultant system of exhaustive and exclusive hypotheses is expansive (continuous) and presents an overwhelming complex assessment task in cases where general specification of the likelihood function for \( P(H|E) \) is situation dependent, as is the case in all substantive audit-testing environments. Though not a problem in principle for auditors relying on the Bayesian method, such likelihood function assessment is an unresolved and difficult practical problem to which considerable auditing research efforts have been devoted (Kaplan 1973 and Loebbecke and Neter 1976). At best the auditor relies on estimates or approximations of probabilities of the form \( P(E|H) \), and hence of probabilities of the form \( P(E) \), since probabilities of this second form are computed as marginal probabilities by the formula \( P(E) = \sum P(H) \cdot P(E|H) \).

Let us suppose, however, that the auditor can assign probabilities in accordance with the subjectivist Bayesian view, and that the auditor can assess the proper loss function. Such a supposition would allow the auditor to proceed with successful completion of the Bayesian inference process. However, even with these (sizable) assumptions, the Bayesian method still faces an insurmountable difficulty. In particular, for the auditor interested in belief formation, the result of the Bayesian inference process is in principle unsatisfactory because the subjectivist Bayesian method overtly couples inference with decision-making rather than opinion or belief formation.16

**The Vexing Characteristics of the Bayesian Foundation**

The fundamental criticism of the Bayesian method of attestation is its inability to lead to professional opinions (that is, reporting judgments). The auditor requires an inference process that leads to a judgment about the fairness of the financial statements under investigation. The subjectivist Bayesian methodology, however, is overtly concerned with making decisions, not with making judgments.
The Bayesian methodology provides a technique for selecting an audit report, the content of which is (inductively) independent of the auditor's beliefs about the fairness of the financial statements under investigation. It relies upon a behavioralist inference process which ends in the assignment of personal probabilities to hypotheses, and a decision process which incorporates those probabilities in the specification of an expected loss function. Further, via the Bernoulli principle, the subjectivist Bayesian method leads directly from the assignment of personal probabilities to hypotheses, then to the selection of an optimal report. But such a report cannot reveal the auditor's judgment about the fairness of the financial statements under investigation, because the inference process yielded no such judgment. Thus, the Bayesian method of attestation imposes a vexing nonstandard interpretation on the concept of an "opinion."

An appropriate report issued by an auditor who has employed the Bayesian method of attestation would read "on the basis of my investigation, and while I have reached no opinion and formulated no judgment, I have concluded that my own best interest is served by stating that the financial statements which I examined fairly present . . . ." Recall that the standard report currently required of the auditor, states "on the basis of my investigation, it is my opinion that the financial statements which I examined fairly present . . . . (and so on)." The most significant difference between the two report types is that the former (Bayesian) report does not contain any revealed judgment while such a judgment is the focus of the latter report.

A defense of the Bayesian method of attestation must render the hypothetical report statement reasonable. Scott (1977), for example, attempts such a defense by giving a major role to the connection between multiuser needs and the auditor's problem. In particular, Scott (1973, p. 305) poses the auditor's problem as follows: "... the auditor is engaged by the firm to report whether, in his opinion, that firm's financial statements fairly present its financial position and results of operations." If "opinion" is understood to mean something like "a belief not based on absolute certainty," then the Bayesian method of attestation is incapable of leading to a resolution of the auditor's problem. The Bayesian method of attestation does not lead to any cognitive artifact which could be construed as a judgment; instead it provides probability assignments via inference, and an action (report) via decision. But, the report rendered via the Bayesian method of attestation actually does not reveal the auditor's beliefs (opinion) since, under the rules of the subjectivist Bayesian methodology, none have been formulated. The Bayesian method of attesta-
tion, thus, imposes its own interpretation upon the concept designated as "opinion," and that interpretation does not coincide with the usual interpretation of the "professional opinion." Further, it is inconsistent with the interpretation implicit in the statement of the auditor's problem imposed by Scott's formulation of the Bayesian method of attestation.

There is, however, at least one class of decision problems for which the Bayesian method of inference and action choice is appropriate. This is the class in which an economically optimal decision and not a professional opinion is the appropriate artifact of analysis. Specifically, the Bayesian method properly constitutes a computation method for determining the appropriate risk component of the auditor's fee. The cost of an audit may be viewed as consisting of two components — the expense of conducting the audit, and the expense of bearing the risk for issuance of an inaccurate audit report. The Bayesian method provides a proper method for computing the latter component of audit cost. However, the formulation and resolution of this problem — the auditor's insurance problem — remains an open issue.

EPISTEMIC METHODS OF ATTESTATION

This section considers three epistemic methods of attestation, each of which avoids the problems of inference and decision faced by the traditional and Bayesian methods of attestation. Each epistemic method of attestation consists of two processes: inference, which governs the formulation of beliefs about the fairness of the presentation of financial statements, and decision, during which the auditor's report is selected. Each epistemic method considered employs a distinct inference process; but all of the epistemic methods employ an informal decision process, requiring the auditor to report those conclusions about fairness which were developed via the inference process. That is, each method requires the auditor to be truthful in composing the audit report (see Table).

The first two epistemic methods of attestation considered herein employ an inference process based upon the Bernoulli principle of maximizing expected (epistemic) utility. Under these methods, the auditor's inference problem is viewed as one of epistemological decision making. That is, on the basis of evidence, the auditor is to select from a class of hypotheses one hypothesis considered worthy of belief. There are as many acts available to the auditor as there are hypotheses, and the \( i \)th act consists of accepting (as worthy of belief) the \( i \)th hypothesis, and thereby automatically rejecting all other
hypotheses. The outcome of accepting a true hypothesis is the correct expansion of the auditor's knowledge base, while acceptance of a false hypothesis results in incorrect expansion of that base. The notion of utility which applies to these outcomes was first explicitly introduced by Hempel as follows:

This much is clear; the utilities should reflect the value or disvalue which the different outcomes have from the point of view of pure scientific research rather than the practical advantages or disadvantages that might result from the application of an accepted hypothesis, according as the latter is true or false. Let me refer to the kind of utilities thus vaguely characterized as purely scientific, or epistemic, utilities. (1960, p. 465)

Let \( u(h, t) \) and \( u(h, f) \) be the epistemic utility of accepting hypothesis \( h \) when it is true and false, respectively. Then on the basis of evidence \( e \), the expected epistemic utility of accepting \( h \) on the basis of \( e \) is

\[
u(h | e) = P(h | e) \cdot u(h, t) + P(\sim h | e) \cdot u(h, f), \]

where \( \sim \) denotes negation. An epistemic method of attestation based upon the Bernoulli principle employs a rule of inference of the form, accept that \( h \) which maximizes \( \nu(h | e) \). There are as many decision-theoretic epistemic methods of attestation as there are specifications or \( u(\cdot, t) \) and \( u(\cdot, f) \). We consider two such methods: (1) a method of satisficing reported in Marschak (1974) and (2) a method based on a measure of logical content due to Hintikka and Pietarinen (1966). The remaining epistemic method of attestation is based upon an adequacy criterion.

The inference process of a decision-theoretic epistemic method of attestation (that is, a method based upon the Bernoulli principle of maximizing expected epistemic utility) responds to the question: on the basis of evidence \( e \), which hypothesis is most worthy of belief? Alternatively, an inference process based on an adequacy criterion responds to the question: on the basis of evidence \( e \), which hypotheses are worthy of belief? Note that the answer to the former question is the specification of one particular hypothesis, whereas the answer to the latter can involve any number of hypotheses, including zero. An epistemic adequacy criterion employs a rule of inference of the form, accept those hypotheses \( h \) such that \( \Gamma(h | e) > 1 - \epsilon \), \( 0 < \epsilon \leq 0.5 \), where \( \Gamma \) is an (information-theoretic) measure.

**Epistemic Methods of Attestation Based upon Decision-Theoretic Criteria**

Epistemic decision-theoretic methods of attestation presuppose that the auditor must select one hypothesis concerning the magnitude of error present in the financial statements under investigation. Let \( H \)
Let \( h = \{h_1, \ldots, h_n\} \) be the class of hypotheses under consideration, and let \( e \) be the evidential matter collected by the auditor. Then the auditor selects as optimal that \( h \in H \) for which \( u(h|e) = P(h|e)u(h,t) + P(\sim h|e)u(h,f) \) is maximized. If the class \( H \) constitutes a fine partition, then there may be numerous hypotheses which are equivalent to the proposition, “the financial statements are fair,” and likewise numerous hypotheses equivalent to the proposition, “the financial statements are materially in error.” If the auditor accepts a hypothesis of the latter kind, he/she can seek an adjustment. The class \( H \) may contain only two hypotheses: (1) “the financial statement is fair,” and (2) “the financial statement is materially in error.” Note, however, that an epistemic method does not require, as does the traditional method, that \( H \) have only two elements.

The first decision-theoretic rule of epistemic inference is due to Marschak and employs a satisfying utility function \( u \) where \( u(h,t) = r \) and \( u(h,f) = s \) for all \( h \in H \), with \( r > s \). The affine transformation \( u'(h,\cdot) = [1/(r-s)][u(h,\cdot)] - \frac{s}{r-s} \) induces the simple utility function \( u'(h,t) = 1 \) and \( u'(h,f) = 0 \), for all \( h \in H \). Then the expected epistemic utility function is \( u(h|e) = P(h|e) \cdot (1) + P(\sim h|e) \cdot (0) = P(h|e) \) and yields the acceptance rule: accept \( h \in H \) on the basis of \( e \) so as to maximize \( u(h|e) = P(h|e) \).

This rule of acceptance is very sensitive to the assignment of prior probabilities and, under simple conditions, it requires that the auditor ignore all of the evidential matter in forming a belief. For ease of exposition let \( H = \{h, \sim h\} \), that is, the specification of \( H \) required by the traditional method of attestation. And, let \( e_1 \) and \( e_2 \) be two potential evidence statements where \( e_1 \) is indicative of \( h \), and \( e_2 \) of \( \sim h \).

The hypothesis \( h \) will be accepted on the basis of \( e_i \) if and only if

\[
(1) \quad P(h|e_i) > P(\sim h|e_i),
\]

that is, if and only if

\[
(1a) \quad \frac{P(h)}{P(\sim h)} > \frac{P(e_i|\sim h)}{P(e_i|h)}.
\]

Similarly, \( \sim h \) will be accepted on the basis of \( e_2 \) if and only if

\[
(2) \quad P(h|e_2) < P(\sim h|e_2),
\]

that is, if and only if

\[
(2a) \quad \frac{P(h)}{P(\sim h)} < \frac{P(e_2|\sim h)}{P(e_2|h)}.
\]
Furthermore,

\[
\frac{P(e_1 \mid \neg h)}{P(e_1 \mid h)} < \frac{P(e_2 \mid \neg h)}{P(e_2 \mid h)}.
\]

The simplex \( P(h) + P(\neg h) = 1 \) is therefore segmented into three regions as shown by the Chart, where

Region I = \{ \langle P(\neg h), P(h) \rangle > | 0 \leq \frac{P(h)}{P(\neg h)} < \frac{P(e_1 \mid \neg h)}{P(e_1 \mid h)} \},

Region II = \{ \langle P(\neg h), P(h) \rangle > | \frac{P(e_1 \mid \neg h)}{P(e_1 \mid h)} \leq \frac{P(h)}{P(\neg h)} \leq \frac{P(e_2 \mid \neg h)}{P(e_2 \mid h)} \},

Region III = \{ \langle P(\neg h), P(h) \rangle > | \frac{P(e_2 \mid \neg h)}{P(e_2 \mid h)} < \frac{P(h)}{P(\neg h)} \leq \infty \}.

The condition which determines Region I is equivalent to \( P(e_1 \mid h) \)

Chart. Epistemic Inference with Decision-Theoretic Criteria
Accordingly, the prior probability pair \(<P(\sim h), P(h)>\) falls in Region I only if \(h\) is sufficiently lower in probability than \(\sim h\) so that, given \(e_1\), \(P(e_1|h) < P(e_1|\sim h)\). Conversely, the condition which determines Region III is equivalent to \(P(e_2|\sim h) < P(e_2|h)\) and implies \(P(\sim h) < P(h)\). Thus the pair \(<P(\sim h), P(h)>\) falls in Region III only if \(\sim h\) is sufficiently lower in probability than \(h\) so that, given \(e_2\), \(P(e_2|\sim h) < P(e_2|h)\). The Marschak rule, therefore, admits of unusual inference behavior only when one of \(h\) and \(\sim h\) has a very high prior probability. That is, the prior probability assignment has unusual inference in Regions I and III. Contrariwise, the condition which determines Region II is equivalent to

\[
\frac{P(e_1|\sim h)}{P(e_1|h)} \leq \frac{P(e_2|\sim h)}{P(e_2|h)},
\]

which is natural given the presumption that \(e_1\) is indicative of \(h\) and \(e_2\) of \(\sim h\). That is, in Region II the prior probabilities are more or less equal, and therefore are not capable of overwhelming the evidence.24

The influence of prior probabilities detailed above can be mitigated if the auditor employs a process generating highly reliable audit evidence. Note that \(P(e_1|h)\) and \(P(e_2|\sim h)\) are the reliability probabilities of the audit evidence process; that is, the probabilities that the audit process will engender the signal \(e_1\) \((e_2)\) when in fact the financial statements under investigation fairly present (are materially in error). Note, that as the reliability of the audit evidence process increases, \(P(e_1|h) \to 1\) and \(P(e_2|\sim h) \to 1\), whereupon \([P(e_1|\sim h)]/[P(e_1|h)] \to 0\) and \([P(e_2|\sim h)]/[P(e_2|h)] \to \infty\), so that Regions I and III vanish. Therefore, if the auditor employs an evidence process which would be a perfect predictor, the unusual behavior induced by the Marschak rule is avoided. Of course, auditors do not possess such a predictor.

Hintikka and Pietarinen (1966) offer an epistemic utility function which is based on a classical measure of information content. The logical content of a sentence is the class of all sentences that it (logically) implies (Carnap 1963, pp. 405–69). A theorem of the probability calculus requires that if \(h_1\) logically implies \(h_2\), then \(P(h_1) < P(h_2)\).25 Thus the size of the logical content of a sentence and the prior probability of that sentence vary inversely. A metric for logical content is \(\text{cont}(h) = 1 - P(h)\), a simple inverse function of \(P(h)\). Now define \(u(h,t)\) and \(u(h,f)\) as follows: \(u(h,t) = \text{cont}(h)\), and \(u(h,f) = \text{cont}(\sim h)\).26 Then, given evidence \(e\), the expected epistemic utility function is \(u(h|e) = P(h|e) - P(h)\).27
A rule of acceptance based upon the Hintikka-Pietarinen utility specification does not admit the unusual inference behavior exhibited by the Marschak rule. Consider again the hypotheses $h$ and $\sim h$, and the evidence statements $e_1$ and $e_2$. The hypothesis $h$ is accepted over $\sim h$ on the basis of $e_1$ if and only if

$$P(h \mid e_1) - P(h) > P(\sim h \mid e_1) - P(\sim h),$$

that is, if and only if

$$\frac{P(h)}{P(\sim h)} > \frac{P(e_1 \mid \sim h) - P(e_1)}{P(e_1 \mid h) - P(e_1)}.$$

(4a)

Similarly, $\sim h$ is accepted over $h$ on the basis of $e_2$ if and only if

$$P(\sim h \mid e_2) - P(\sim h) > P(h \mid e_2) - P(h),$$

that is, if and only if

$$\frac{P(h)}{P(\sim h)} < \frac{[P(e_2 \mid \sim h) - P(e_2)]}{[P(e_2 \mid h) - P(e_2)]}.$$

(5a)

Given that $\{e_1, e_2\}$ is an exhaustive and exclusive partition of the evidence set, then

$$\frac{[P(e_1 \mid \sim h) - P(e_1)]}{[P(e_1 \mid h) - P(e_1)]} = \frac{[P(e_2 \mid \sim h) - P(e_2)]}{[P(e_2 \mid h) - P(e_2)]},$$

so there are no regions of the simplex $P(h) + P(\sim h) = 1$, wherein either $h$ or $\sim h$ is always accepted regardless of the evidence. Thus, the Hintikka-Pietarinen rule of acceptance requires that the evidence always be employed in accepting a hypothesis.

An Epistemic Method of Attestation
Based upon an Adequacy Criterion

The Marschak rule of acceptance can be converted into the adequacy rule: Accept $h \in H$ on the basis of evidence $e$ if $P(h \mid e) > 1 - \epsilon$, and only if $P(h \mid e) \geq 1 - \epsilon$, for $0 < \epsilon \leq 0.5$. This is the classical rule of high probability, which reflects inductive common sense.

The conversion is made as follows: As before, let $u(h, t) = 1$, $u(h, f) = 0$, so that the expected utility function $\nu$ yields $\nu(h \mid e) = P(h \mid e)$ and $\nu(\sim h \mid e) = 1 - P(h \mid e)$. Now let $s$ denote the (expected) utility of suspending judgment, that is, $s(\emptyset \mid e) = s$. Then the Bernoulli principle of maximizing expected utility yields:

Accept $h$ on the basis of $e$ if $P(h \mid e) > s$. If $P(h \mid e) = s$, then $h$ may be accepted or left in suspense. If $1 - s < P(h \mid e) < s$, then suspend judgment. (Hilpinen 1968, p. 87)
Rewriting $s$ as $1 - \epsilon$ yields the high probability rule of acceptance: Accept $h$ on the basis of $e$ if $P(h|e) > 1 - \epsilon$, and only if $P(h|e) \geq 1 - \epsilon$.

Note that the value $s$ serves as a lower bound on the expected utility function $u(\cdot|e)$ such that if, for any $h$, $u(h|e) < s$, then $h$ is considered inadequate and not accepted. The introduction of $s$ thereby converts a decision-theoretic rule of acceptance into an adequacy rule.

The rule of high probability, like the Marschak rule, results in unusual inference behavior. Consider again $h$ and $\sim h$ and their acceptance on the basis of evidence sentences $e_1$ and $e_2$, respectively. By the rule of high probability $h$ will be accepted on the basis of $e_1$ if and only if

$$P(h) > \frac{(1-\epsilon)P(e_1)}{P(e_1|h)}.$$  

(7)

Similarly, $\sim h$ will be accepted on the basis of $e_2$ if and only if

$$P(\sim h| e_2) > 1 - \epsilon,$$

that is, if and only if

$$P(\sim h) > \frac{(1-\epsilon)P(e_2)}{P(e_2|\sim h)}.$$  

(8a)

Thus, there are lower bounds for the prior probabilities $P(h)$ and $P(\sim h)$, such that, if one of these bounds is exceeded, either $h$ or $\sim h$ will be accepted regardless of the evidence. Further, the rule of high probability is conservative in that it precludes conjunctive closure of knowledge.²¹

**The Decision Process for Epistemic Methods of Attestation**

As noted earlier, an epistemic method of attestation consists of two processes — inference and decision. The decision process relates the beliefs arrived at via the inference process to the selection of an audit report. If the auditor infers that the financial statements under investigation are materially accurate, he/she simply reveals that belief in his report. Similarly, if the auditor infers that the statements are materially inaccurate, he/she either seeks an adjustment or reveals that belief. The choice between the latter two actions is a matter of judgment. Finally, the auditor requires additional evidence under one of two conditions: (1) when a decision-theoretic inference process yields a tie between (or among) a number of best hypotheses and (2) when an adequacy criterion inference process either suggests that all hypotheses are unacceptable, or that conflicting hypotheses are ac-
ceptable. The auditor’s search terminates when the tie is broken, when at least one hypothesis is accepted, or when the conflict is resolved.

The decision process for an epistemic method of attestation is predicated upon the view that the auditor’s beliefs will be truthfully revealed. This view is demanded by the ethical standards to which the auditor is committed. If, however, the auditor should choose to lie, then the decision process of the epistemic methods of attestation effectively would be replaced by the expected utility framework.

CONCLUSIONS

The most commonly practiced method of attestation is the traditional method, or variations thereof, and the most strongly supported alternative method is the Bayesian method. This paper has shown that neither method is an adequate means for resolving the auditor’s attestation problem. The traditional method is inadequate because it cannot yield judgments about a particular account or set of financial statements. The Bayesian method is inadequate because it is incapable of providing judgments. We posit that methods of attestation that avoid these inadequacies belong to the class of epistemic methods of attestation and have considered three such methods.

The epistemic method based upon the Hintikka-Pietarinen acceptance rule is especially well suited for resolving the auditor’s problem. It provides for the formulation of a judgment about the fairness of a particular financial statement’s presentation, and governs the selection of an audit report that is consistent with the inferred judgment. The inference process of this method employs an epistemic utility function, which values a hypothesis according to its logical content, so that the expected (epistemic) utility of a particular hypothesis \( h \) about the fairness of a financial statement presentation given evidential matter \( e \) is \( u(h|e) = P(h|e) - P(h) \). The hypothesis which maximizes the expected epistemic utility function \( u(h|e) \) is accepted as the hypothesis most worthy of belief and is the auditor’s inferred judgment.

The decision process employed in each epistemic method requires that the auditor select an opinion in accordance with the results of the inference process. To interpret “in accordance with” to mean that the auditor must select an opinion which literally asserts his inferred judgment is naive, and leads to counterintuitive decision behavior. For example, if the auditor considers the two hypotheses “the financial statement is fair” and “the financial statement is unfair,” and accepts,
say, the former by a very small margin over the latter, and naively employs the decision process, then he/she must select a "clean" opinion. In such a situation, however, the auditor may prefer to issue a report containing something other than an unqualified opinion. Second, any attempt to resolve the auditor's attestation problem involves at least part of the more general problem of induction. The latter has neither an agreed-upon statement nor a solution. Indeed, the central conclusion of this paper is that the auditor's attestation problem is deeply embedded in the more general problem of induction. Any attempt to resolve the auditor's problem which overlooks this connection will be inadequate.

Finally, the auditor's problem, as considered here, presumes that the audit is performed by one person. In reality, audits are conducted by teams of investigators. The "auditor" in such cases is a collective, not a single individual. Thus, a complete statement of the auditor's attestation problem must acknowledge the structure of the audit team. Resolution of this larger variant of the auditor's problem awaits the development of a theory of team inference. Such a theory might parallel the economic theory of teams (Marschak and Radner 1974). However, the Marschak-Radner theory is not itself a theory of team inference for exactly the same reason that the Bayesian method of attestation fails to be a method of inference for the auditor. Simply put, the single-person audit problem is embedded in the traditional problem of induction; the team-audit problem is embedded in a more general problem of induction. Interestingly, the latter problem of induction has not received any attention. Thus, the problem of team induction and the team-audit problem remain open issues in both philosophy and accounting.

NOTES

*We wish to acknowledge the many helpful comments and suggestions made by a number of individuals on earlier drafts of this paper. Especially helpful have been the remarks of W. Bente, B. Cushing, R. Hamilton, J. Kennelly, and W. Scott.


2. Throughout this paper "materially accurate," "faithful," and "fair" are used synonymously.

3. For detailed presentations of the cognitivist and behavioralist positions see Hilpinen (1968) and Carnap (1963), respectively. For a discussion of the two views see Niiniluoto and Tuomela (1978).

4. This table also provides, for later reference, a similar profile of the
two major classes of epistemic methods of inference and decision discussed in later portions of this paper.

5. Kyburg notes that "...there is no way of combining the two kinds of error to get an overall 'probability' of error. Furthermore, these two kinds of error are qualitatively different. It is thus a blunder from the very beginning even to attempt to combine them into a single blanket 'error' as we might be tempted to do, and as Bayesian theorists, for example, do." (1974, p. 33)

6. On the convention of classical statistics see, for example, Kyburg (1974, p. 34f).


8. Neyman (1950) argues that the Neyman-Pearson theory is not a cognitivist theory; he asserts that the term "inductive inference" is misleading and should be replaced by "inductive behavior." Contrairewise, Birnbaum (1977) argues that the Neyman-Pearson theory admits of both cognitivist and behaviorist interpretations. See, for example, Kyburg (1974, pp. 23-24), and Birnbaum (1977, pp. 26-33).

9. The justification of the Neyman-Pearson inference methodology lies in the long-run properties of the method. See, for example, Kyburg (1974, pp. 50-59).

10. See for example, Elliott and Rogers (1972, p. 46), as earlier quoted.

11. Subjectivist Bayesian methodology should not be confused with empirical Bayesian methodology. On the latter see Robbins and Samuel (1962) and Sennetti and Kakar (1975). Further, recall that the traditional method of attestation is an adaptation, rather than a direct application, of the Neyman-Pearson methodology.

12. On the notion and importance of "realized events" see Marschak (1975).

13. The penalty for holding incoherent degrees of belief is the absorption of a Dutch book — a system of bets wherein regardless of the realized event or events, the individual loses. The notion of a Dutch book is due to Kyburg and the notion of coherence is due to de Fennetti. On both notions see Kyburg (1974, p. 95).

14. To avoid ruin, the Bayesian auditor may wish to adopt a strategy of lying. On gambler's ruin see Moran (1968); on the history of gambler's ruin see Thatcher (1957).

15. Note however that for the Neyman-Pearson auditor, the specification of \( P(E|H) \) for nonsampling evidence is a problem in principle. The classical mode of inference has no mechanism for encompassing data-generating functions that cannot be studied on the basis of long-run frequencies of occurrence.

16. For much the same reason, the underlying subjectivist Bayesian inference process is incapable, in theory, of supplying all of the information required by the Bayesian decision process. For an extended discussion of the incapacity of the behaviorist inference process to provide the outcome mapping required, see Dacey (forthcoming and 1976). The point is made indirectly in Sellars (1946). Briefly, the argument runs as follows: In order to support an adequate method of attestation the subjectivist Bayesian
inference process must provide all of the information required by the Bayesian decision process. However, the Bayesian inference process, like all behavioralist inference processes, cannot provide a knowledge of the outcomes resulting from the performance of each action across the various states of nature. The Bayesian inference process is incapable of providing the outcome mapping presupposed by the Bayesian decision process. The standard formulation of the Bayesian decision process involves a problem with state space $\chi$ and action space $A$, and a decision maker with probability measure $p$ defined over an algebra on $\chi$, an outcome mapping $z$ from $\chi \times A$ to the outcome space $O$, and loss function $l$ from $O$ to the real numbers. The Bernoulli principle then asserts that the best action in $A$ is that which minimizes expected loss, that is, the individual behaves optimally by selecting $a \in A$ so as to minimize $L(a) = \sum_{\chi} p(\chi) l(z(\chi, a))$. The outcome mapping $z$ is equivalent to a statement of the form "for all $\chi, a \in \chi \times A$, if action $a$ is performed in state $\chi$, then outcome $z(\chi, a)$ will occur." Such a statement must be known to be true (or accepted as true) by the decision maker if he is to employ the Bernoulli principle. The subjectivist Bayesian inference process, however, provides only probability statements. Thus the outcome mapping, which is not a probability statement, cannot be provided by the subjectivist Bayesian inference process. The inability of subjectivist Bayesian inference per se to provide all of the information required by the subjectivist Bayesian decision process is a problem in principle. The outcome mapping employed within a Bayesian decision process must come from a source external to the Bayesian inference process. The manifestation of this problem in the auditing context constitutes a problem in practice. (See, for example, Moriarity and Barron (1976) and Ward (1976 and 1976).) Some recent studies indicate that the problem of determining the proper utility structure in a multiple-user setting (in accounting and auditing) may not be without resolution. See Cushing (1976) and Scott (1975). The examination of a firm's financial statements and supporting records alone cannot possibly provide the auditor with a knowledge of the outcomes resulting from issuing each of the alternative reports in each of the various states. In order to make the Bayesian method of attestation practicable, the auditor must employ a cognitivist inference process, which by design is capable of generating the required outcome mapping. But this constitutes an ad hoc effort at patchwork, which, as we argue below, is avoidable if the auditor employs certain epistemic methods for attestation.

17. The introduction of user preference orderings as a means of allowing the Bayesian auditor to optimize a variety of user needs also introduces the standard paradoxes of social choice theory into auditing. Scott points out that, under certain conditions, such paradoxes can be resolved. If all of those who value information presented by audited financial statements possess single-peaked utility structures, then group (societal) preference orderings can be obtained. Even so, resolution of the auditor's problem cannot be obtained without combining user utility functions. To do so the auditor must have available individual user utility functions whose metrics are additive, and the auditor must be able to assign relative weights to the utility functions of all users. The former task could, in practice, be insurmountable if user preferences involve noncommensurable properties (for example, social progress, quality of life, environmental impact, and so on,
as opposed to strictly economic properties). The problem of weighting user utility functions immediately reinvokes the paradoxes of social theory. Note as well, that while single-peakedness is intuitively appealing in some circumstances it is, at best, without empirical support and counterintuitive with respect to hypothetical users who prefer income to loss but would prefer, once a loss becomes significant, larger deficits (the big bath) to smaller ones.

18. Mautz and Sharaf (1961) constitute an extensive explication of the concept of a "professional opinion." Their notion of a professional opinion is consistent only with a cognitivist inference methodology, since they specify the auditor's goal to be the elimination of the probability of error while admitting the possibility of error. See, for example, Mautz and Sharaf (1961, p. 47).

19. Recall that the traditional method of attestation cannot contribute to a resolution of the auditor's insurance problem, since it places the risk of an erroneous audit report on the wrong societal unit.

20. The concept of "epistemic utility" can be traced to C. S. Pierce, a point we owe to I. J. Good (private communication). Incidentally, Good suggests the term "quasi-utility," in lieu of "epistemic utility," to draw attention to the distinction between practical action and mental (or belief formation) action. Note that, as described by Hempel, an epistemic utility is very unlike the concept of utility employed in the subjectivist Bayesian methodology of decision. The latter assigns numerical values to realizables (that is, economic) events, but the former assigns numerical values to mental events, that is, the auditor's knowledge about the true state of the firm under investigation. It would appear that Hempel's "scientific utilities" capture the preferences of Scott's "scientific report" (Scott 1977, p. 124).

21. Herein lies a major distinction between decision-theoretic and adequacy methods of acceptance. Unless there is a tie for best hypothesis, a decision-theoretic method will automatically select a unique best hypothesis, and it will always select at least one hypothesis. Contrariwise, an adequacy method may select all of the hypotheses under consideration, some of them, or none of them. Note that if a decision-theoretic method yields a tie or if an adequacy method selects no hypothesis, then the auditor must conclude that further evidence is required, and must go about collecting the same.

22. Levi (1961) considers this utility function in an early discussion of the relevance of decision theory to philosophy and dismisses it as being irrelevant. The function is attributed here to Marschak because he argues for its relevance and practicability. On the general relevance of decision theory to philosophy also see Suppes (1961).

23. Definition: evidence $e_1$ is indicative of $h$ if and only if $P(e_1 | h) > P(e_1 | \neg h)$. Similarly, evidence $e_2$ is indicative of $\neg h$ if and only if $P(e_2 | \neg h) > P(e_2 | h)$.

24. If the auditor's initial probabilities of $h$ and $\neg h$ fall in Region I, then he/she will accept $\neg h$ regardless of the evidence. If the auditor's initial probabilities fall in III, then he/she will accept $h$ regardless of the evidence. Only if the auditor's initial probabilities fall in II will the auditor base acceptance of either $h$ or $\neg h$ on the evidence. Note that the auditor's susceptibility to this unusual inference behavior places him/her in a position to be manipulated by the client. For example, if $h$ is the hypothesis that the financial statements are fair, and if the client can induce the auditor to hold initial probabilities which place the latter in I, then the client has manipulated
the auditor into rendering a clean opinion, regardless of the evidence accumulated by the auditor or the true state of the financial statement. The mechanics of manipulation are detailed in Dacey et al. (1977).

25. Note that \( P(h) \leq P(h_b) \) does not imply that \( h \) logically implies \( h_b \). Thus, the derivation of the Hintikka-Pietarinen rule is not based upon an equivalence between prior probability and logical content. This represents a weakness in the derivation of the rule, a point we owe to Marschak (private communication).

26. The rationale for this specification of \( u \) is as follows: “Let us consider a theoretician (that is, a person whose only relevant utility is information) who decides to accept a hypothesis \( h \). This hypothesis has a certain information-content \( \text{cont}(h) \). If \( h \) is true, the utility of his decision is the valid information he has gained, that is, \( \text{cont}(h) \). If \( h \) is false, it is natural to say that his disutility or loss is measured by the information he has lost because of his wrong choice between \( h \) and \( \sim h \), that is, by the information he would have gained if he had accepted \( \sim h \) instead of \( h \). This is measured by \( \text{cont}(\sim h) \), where the utility in this case is \( -\text{cont}(\sim h) \)” (Hintikka and Pietarinen 1966, p. 107–08).

27. This specification of \( u \) has been proposed by Levi (1967) as a measure of the explanatory power of the hypothesis \( h \).

28. The proof is as follows:

\[
\frac{P(e_1 | \sim h) - P(e_1)}{P(e_1 | h) - P(e_1)} = \frac{(1 - P(e_1 | \sim h)) - (1 - P(e_1))}{(1 - P(e_1 | h)) - (1 - P(e_1))} = \frac{P(h) - P(e_1 | \sim h)}{P(h) - P(e_1 | h)}
\]

29. The rule has other favorable properties, which are elucidated in Hilpinen (1968), Hintikka and Pietarinen (1966), and Niiniluoto and Tuomela (1973). Some of the shortcomings of the rule are detailed in Lehrer (1974, p. 220–24). A variation of the Hintikka-Pietarinen rule of belief formation, initially due to Finch (1960) and developed by Lehrer (1974), is obtained when \( u(h | e) = \text{cont}(h) / P(h) \), and \( u(h | \sim e) = -\text{cont}(\sim h) / P(h) \). The resulting expected epistemic utility function is \( u(h | e) = P(h | e) - P(h) / P(h) \) which is the Hintikka-Pietarinen expected utility function normed by \( P(h) \). Note also that \( u(h | e) \) can be written as \( u(h | e) = [P(h | e) / P(h)] - 1 \). Thus to maximize \( u(h | e) \) is to maximize \( P(h | e) / P(h) \).

The Finch-Lehrer rule of acceptance does not allow the inefficient processing of evidence. To prove this consider again the hypothesis \( h \) and \( \sim h \), and the evidence statements \( e_1 \) and \( e_2 \). The hypothesis \( h \) is accepted over \( \sim h \) on the basis of \( e_1 \) if and only if \( [P(h | e_1) / P(h)] - 1 > [P(\sim h | e_1) / P(\sim h)] - 1 \), which reduces to \( P(e_1 | h) > P(e_1 | \sim h) \). Similarly, \( \sim h \) is accepted over \( h \) on the basis of \( e_2 \) if and only if \( [P(\sim h | e_2) / P(\sim h)] - 1 > [P(h | e_2) / P(h)] - 1 \), which reduces to \( P(e_2 | \sim h) > P(e_2 | h) \). But the conditions which determine the acceptance of \( h \) and \( \sim h \) (given \( e_1 \) and \( e_2 \), respectively) are just the definitions that \( e_1 \) indicates \( h \) and \( e_2 \) indicates \( \sim h \), which was presumed at the outset of the proof. Thus there are no segments of the
line $P(h) + P(\neg h) = 1$ wherein $h$ or $\neg h$ is always accepted regardless of the evidence.

Thus both the Hintikka-Pietarinen and Finch-Lehrer rules of acceptance are superior to the Marschak rule, and cannot, on the basis of unusual inference behavior, be distinguished one from another. However, Lehrer argues that the Finch-Lehrer rule is superior to the Hintikka-Pietarinen rule on the grounds (Lehrer 1974, p. 227f).

30. See Hilpinen (1968) for an extended discussion of this rule. Acceptance on the basis of high probability reflects common sense: "The customary conception of inductive inference is that sufficiently strong inductive support makes a hypothesis worthy of belief. If the degree of inductive support is explicated in terms of probability, it is natural to assume that a high probability is a necessary and sufficient condition of acceptability" (p. 24). See also Cramer (1946, p. 150).

Furthermore, Mautz and Sharaf (1961, p. 72f) can be interpreted as advancing the high probability rule as the proper acceptance rule for auditing.

31. This result is due to Kyburg (1970 and 1961). Consider the following three principles:

The weak deduction principle. If $S$ is a body of accepted statements, and $S_1$ belongs to $S$ and $S$ entails (that is, logically implies) $S_2$, then $S_2$ belongs to $S$.

The weak consistency principle. If $S$ is a body of accepted statements, then there is no member of $S$ that is self-contradictory (that is, $S_i$ & $\neg S_i$ is not a member of $S$).

The conjunctive closure principle. If $S$ is a body of accepted statements and $S_1$ belongs to $S$ and $S_2$ belongs to $S$, then $S_1$ & $S_2$ belong to $S$.

These principles reflect basic intuition and commonsense vis-à-vis acceptance, induction, and belief. They are, however, inconsistent and imply the lottery paradox. The lottery paradox is as follows: Consider a lottery with exactly one winner and one million tickets, each with the same chance of winning. The hypothesis $h_i = \text{"the } i\text{th ticket will not win}"$ has a very high probability, that is, $1 - (1/10)^{10}$, and is thus acceptable on the high probability rule. This is true for $i = 1, 2, \ldots, 10^9$. Thus, $h_1, \ldots, h_{10^9}$ are accepted. If we conjoin them we are left with the equivalent of "no one will win the lottery." This is inconsistent with the initial description of the lottery, hence the paradox. The principle of conjunctive closure must be abandoned. Kyburg argues, since the principle of weak deduction and weak consistency are so obviously less objectionable. Note that the Marschak rule of belief formation avoids the lottery paradox. On the Marschak rule one would accept the disjunction $h_1 \lor h_2 \lor \ldots \lor h_{10^9}$, which is the logical equivalent of "some ticket will win."

The rule of high probability can be protected from the lottery paradox. The means of protection are due to Lehrer, and are discussed in Lehrer (1974 and 1974), and Dacey (forthcoming). The traditional method of attestation, via its Neyman-Pearson foundation, is also open to the lottery paradox. However, it is not clear how to protect that method from the paradox. On this point see Hilpinen (1968, pp. 46–49).
REFERENCES


On the Fundamental Nature of Professional Opinions: The Traditional, Bayesian, and Epistemic Methods of Attestation: Discussion*

Dacey and Ward's paper begins with two basic premises. The first is that the objective of the auditor who uses classical hypothesis testing is to make reasonably certain that the current year's financial statements do not contain a material error. The second is that the objective of the auditor who employs a Bayesian approach is to choose an optimal act, that is, to choose an audit report to minimize expected loss. These objectives are supported by reference to the literature, specifically to Elliott and Rogers (1972) for the classical approach and to Scott (1973) for the Bayesian approach.

If these premises are accepted, then I find the paper to be a precisely and carefully reasoned document. In particular, it provides some interesting insights into what the auditor's actual behavior is when he/she uses a classical statistical or a decision theoretic approach. I think, however, that several of the premises may be "straw men" or, at least, not broad enough to capture what the auditor really does. Consider the following four comments.
(1) Suppose we redefine the auditor’s objective as the desire to be efficient and effective, to deserve public confidence, or just to survive in a difficult environment. In this broader view, I think that quantitative methods, and statistical methods in particular, have served the auditor reasonably well. They provide a consistent, defensible framework for attacking at least some auditing problems. Of course, because these methods work does not mean that the philosophical basis which underlies them should not be examined.

(2) However, the authors argue that the auditor’s concern, when applying the traditional classical hypothesis-testing methodology, is only with the particular financial statements being audited. This leads to their point that the Neyman-Pearson hypothesis-testing methodology conducted at, say, a β-risk of 5 percent, does not imply that there is a 5 percent risk ex post of material error in this report. Rather, what is implied is an ex ante notion that a long-run reference sequence of repetitions of similar audits would miss a material error 5 percent of the time. Thus, an auditor is not really following the logic of the method when he/she issues a standard report.

However, Dacey and Ward ignore the fact that the auditor will have a portfolio of clients and, indeed, may audit the same client for many years. The audit firm will have a particular audit strategy that, broadly speaking, is applied to all clients. This strategy may well include a specification that statistical tests will run, say, a 5 percent risk of missing a material error. Then, the ex ante reasoning underlying the Neyman-Pearson methodology seems reasonable when applied to a firm-wide audit strategy even though it may not be reasonable, strictly speaking, when applied to an individual audit report.

Also, redundancy must be built into an audit. I doubt that any auditor could survive making audit effectiveness errors 5 percent of the time, for example. Thus, the Neyman-Pearson inference method which is being criticized is not applied in a “pure” scientific fashion in auditing but is supplemented with additional evidence. This raises further questions about the validity of the authors’ assumption that the auditor imposes a well-defined inference model on the choice of an audit report — the formal inference model is only part of what the auditor does.

(3) The Neyman-Pearson and Bayesian methods are separated into different “camps” in the paper. While there are significant philosophical differences between the methods, it would have been useful to point out that Bayesian methods can be interpreted as an extension of classical methods. The vehicle for this interpretation is the likeli-
hood function in Bayes' theorem which, effectively, is the same distribution as that used in classical estimation and hypothesis testing. If the assumed sampling distribution in these tests does not "fit" the data, risks and confidence levels will be affected — see, for example, Teitlebaum and Robinson (1975). If the likelihood function does not fit the data, Bayes' theorem is not a normative vehicle for revision of probability beliefs. In this sense, the two methodologies share many common problems, and I am not sure it is useful to set them up as mutually exclusive.

(4) I disagree with the authors' claim that the Bayesian method of attestation does not involve judgment. Their argument seems to be that, since the auditor's choice of a report is the result of an expected-loss-minimizing calculation, it is not a professional judgment in the usual interpretation. This is not correct. The auditor's judgment is very much a part of the Bayesian methodology. Any application of statistical techniques requires judgment; the question is how this judgment can best be applied. In a Bayesian approach, judgment is applied to assessing prior probabilities and the loss structure. The fact that this judgment is only implicit in the final attestation decision does not alter the fact that it is present. In fact, it seems strange that the authors' express concern about a lack of judgment in the body of the paper; yet in Note 16 they express concern that the auditors' loss function must come from a source external to the inference process — that is, that it requires judgment.

As a result of these considerations, I feel that the authors' premises about audit objectives are too simple for their criticisms about methodology to be very meaningful.

THE EPISTEMIC METHOD OF ATTESTATION

Rejection of the criticisms of classical and Bayesian methodologies does not imply rejection of epistemic methods, however. Therefore, let us consider whether the epistemic approach has the potential to be a useful addition to the auditor's set of quantitative tools. The authors' argument for these methods seems to be based on a rejection of the decision-theoretic notion of the auditor's loss function. Instead of "selfishly" minimizing his/her expected loss, the auditor should rise above these considerations by selecting a report which is best from the standpoint of "society." This looks like a standard of "fairness" and it seems that the authors regard it as the essence of professionalism, at least this is my interpretation of their quotation from Hempel in the fourth section of the paper.
Their attack on the notion of minimizing expected loss ignores the fact that auditor losses can be regarded as originating from financial statement users (Scott 1975). It also flies in the face of much current agency research which regards the auditor, like everyone else, as a rational economic agent (Antle 1982). Nevertheless, for purposes of discussion, let us accept the epistemic notion.

Then, it is no accident that the three epistemic attestation methods suggested depend only on probabilities, not on losses. It is not so much that losses are ignored as that they are cleverly collapsed into the probabilities. The decision rules that are examined can be loosely categorized as (1) accept the hypothesis with the highest posterior probability of being true, (2) accept the hypothesis with the biggest surprise content, and (3) accept those hypotheses with adequate posterior probabilities.

To what extent might these be useful auditor decision rules? One benefit of using them would be an easing of the multiuser problem, which complicates calculation of the Bayesian loss function. This occurs because the epistemic approach requires the auditor to think in broad social terms, such as fairness, rather than specific auditor/user objectives. The question of the extent to which the proposed methods can accomplish this should have received more attention in the paper.

A second potential benefit arises from the complex nature of the audit. As Einhorn and Hogarth (1981) suggest, normative decision theory may break down in an audit context because the environment is too ill-defined or unstable for decision theory concepts to be viable, leading to suggestions that heuristics may be rational and may persist. Here the paper would benefit from reference to the work of Chesley (1983), who also suggests the applicability of an epistemic view of probability to the audit. Chesley suggests that in the real audit environment, the auditor may not be able to specify fully the set of states, acts, and so on, required for a full decision-theoretic approach. Then, the subjective and relative frequency views of probability break down and Chesley argues that the epistemic view provides a replacement. His arguments are not inconsistent with the epistemic decision rules suggested here. Thus, it may be that the epistemic approach provides a vehicle for extending the boundaries of formal analysis in the audit.

There is little discussion in the paper, however, of how the suggested methods might be applied, and what the auditor is really doing in using them. For example, it seems that the high probability rule is sensitive to the fineness of specification of the hypotheses. Suppose
$h = \text{"the statements do not contain a material error,"}$ and $P(h | e)$ is .60, where $e$ is the audit evidence. Then it would be chosen over $P(\sim h | e)$ with a .40 probability. However, if we have three hypotheses: $h_1 = \text{"no error in the statements,"}$ $h_2 = \text{"inmaterial error in the statements,"}$ each with probability of .30, and $h_3 = \text{"material error in the statements,"}$ then $h_3$ would be chosen. Of course, the 3-hypothesis case does not constitute a payoff-relevant partitioning of the state space, but this is just my point. To set up a payoff-relevant set of hypotheses requires consideration of the loss structure, which the authors have previously rejected. Thus, it may not be as easy to dispose of the loss structure as the authors seem to think. I use this example only to illustrate that a lot of questions would arise if the epistemic approach were to be implemented.

In conclusion, to apply an adequacy criterion, I suspend judgment. I think that epistemic methods are worth further investigation. However, the authors will have to justify their approach in a positive fashion rather than by attacking extant methods.

**NOTE**

* The comments of a second discussant were not available for publication.

**REFERENCES**


Session

THREE
Models for Multilocation Audits

If all the statisticians in the world were laid head to toe, they wouldn’t be able to reach a conclusion.\(^1\)

The third standard of field work of the AICPA’s generally accepted auditing standards (GAAS) states that “Sufficient competent evidential matter is to be obtained through inspection, observation, inquiries, and confirmations to afford a reasonable basis for an opinion regarding the financial statements under examination.” The adoption of Statement on Auditing Standards (SAS) 47 (AICPA 1984, “Audit Risk and Materiality in Conducting an Audit”), together with SAS 39 (AICPA 1984, “Audit Sampling”), provides the auditing profession with standards and guidance as to what constitutes sufficient evidential matter. Further, these pronouncements, as a package, provide a fairly comprehensive conceptual and practical framework for the external audit. We expect a flood of literature over the next decade (by both academics and practitioners) analyzing, interpreting, and commenting on this general framework. We hope that some of this literature will address the issues raised in this paper.
A major deficiency of the preceding pronouncements (and of all current auditing texts) is the lack of guidance with respect to determining what constitutes sufficient evidential matter in a multilocation environment. It is rather surprising that this matter has received so little attention when one considers that the problem is encountered on most audits of medium and large entities. For example, determination of the appropriate extent of verification of accounts receivable for an entity with ten locations is straightforward when the records are centralized. However, when the accounting records are decentralized, and it becomes necessary for the auditor to visit a location to carry out the verification procedures, the following two major questions arise: (1) How many locations should be visited? and (2) What extent of the work (sample size) should be carried out at each location visited? The same questions arise when the records are centralized but the physical location of the asset or financial statement item is decentralized. For example, inventory records may be centralized even though the inventory is located at ten warehouses across the country. This circumstance would facilitate a centralized audit of the valuation assertion, but on-site visits would be required to audit the existence assertion.

The problem is not restricted to a single corporate entity in that many audits (particularly in the United States, which does not have a statutory audit requirement like those found in the Canadian and United Kingdom Corporations Acts) involve national and multinational entities which consist of a large number of subsidiary corporations (often in the hundreds). In such cases, it is not uncommon to find the auditor rendering only one opinion — on the consolidated financial position and operations. Here, the auditor is confronted with similar decisions: how many subsidiaries (and which ones) should be audited and how much work should be carried out on the subsidiaries selected in a particular year?

The CICA Extent of Audit Testing research study (1980) reported the results of a “Questionnaire on the Extents and Manner of Testing.” Several questions related to testing in a multilocation environment. From the response, it was clear that there was no consensus within the profession as to whether all locations should be audited every year or whether some form of rotational verification should be used. Several research or discussion papers were prepared by various members of the study group on topics addressed in the study. One of these papers (Leslie 1979) was Phase 1 of the research continued in this paper.

At the present time, most practitioners (whether or not they employ
statistical sampling) determine the number of locations that they visit on a judgmental basis. This approach usually involves visiting the largest locations every year together with a rotating selection from the remainder. Rotational selection results in each location in that category being visited once over a period of years. But, under rotational selection, it can be difficult to evaluate audit findings with respect to the entire population. The locations selected each year may not be representative of those not selected (that is, since the element of randomness is missing when rotational selection is used). Also, the client’s staff probably knows that a rotational selection method is being employed and approximately when to expect a visit. Finally, biases unrelated to audit factors have been known to creep into the selection process (for some reason it seems more important to visit the branch in Florida than the branch in Alaska— in the middle of winter). When an internal audit department exists, such visits are often coordinated with internal audit activities so that external and internal auditors will visit different locations to maximize coverage (of course, with proper planning, the internal auditors will be assigned to branches in Anchorage, Duluth, Grand Forks, Great Falls, and so on, while the external auditors will visit the branches in Las Vegas, San Diego, San Francisco, and Fort Lauderdale).

It probably is safe to conclude that most practitioners base the amount of their work at each location visited on “conventional” sampling extents (engagement materiality and audit risk employed with a statistical sampling plan as if all locations were being visited or, alternatively, a judgmental determination of sample extent in whatever manner “judges” make such decisions). Since we cannot capture the thought process of judges, statistical sampling will be used in the following illustration, taken from Leslie (1979).

The Situation

Business: manufacturer (sales to wholesalers and retailers);
Population: accounts receivable (decentralized records—10 locations equal in size, a simplifying assumption for illustration purposes only);
Book value: $27 million (individual account balances range from $1 to $60,000);
Audit materiality: $300,000;
Substantive sampling risk: 5 percent (after taking into consideration the inherent risk, strength of internal control, effect of analytical review, prior error experience, and so on);
Sampling plan used: dollar-unit sampling; and
Sample size: 270 (the minimum if no errors are expected throughout
the audit — used for simplicity but discussed in the next part of this paper).

When a sample of 270 accounts receivable is selected from the entire population and no errors are found in the confirmation process, the auditor will have achieved the desired objective. If he/she must visit each location because of decentralized records, but decides that this would be too expensive, should he/she still try to achieve the objective? Obviously the answer is yes, since the auditor will still issue the same opinion when fewer than ten locations are visited. Thus, the auditor must decide on the number of locations to be visited and how many accounts he/she will confirm at each location.

If the auditor were to believe that any one branch could contain a material error ($300,000) and that there is more than a remote chance that this fact would go undetected if confirmation procedures were not applied at such branch, the answer is simple — all locations must be visited. While most auditors will find this an unpleasant conclusion, it follows logically from the underlying conditions and cannot be avoided.

Now suppose it is reasonable for the auditor to make an assumption about the maximum amount of error that might go undetected at each branch by the system of internal control and analytical review procedures. (Some auditors might recoil in horror at the thought of making such an assumption, but it should be remembered that auditing requires all kinds of assumptions to be made. In any event, if such an assumption is not reasonable, it will not be possible to achieve the audit objective without performing audit procedures at all branches — something very rarely done by auditors at the present time.) Suppose that in this illustration the auditor were to believe that error aggregating $100,000 or more at any location would be detected, as already mentioned. Then, a material error could be spread over as few as three branches and go undetected if none of the three were selected for confirmation purposes.

The two-stage procedure developed by Leslie (1979) (and discussed further in the next part of this paper) produced a first-stage sample of six out of the ten branches for the $100,000 location error-size limit. If fewer than six locations were visited, the auditor would have already exceeded the 5 percent risk (before sending any confirmations), since the risk of not visiting at least one of the three locations with errors would exceed 5 percent (8.3 percent if five locations were visited and 16.7 percent if four locations were visited).

The next question is, at each location, how many confirmations
should be sent. One approach would be to select 27 at each of the six locations (the same number that would be selected if all locations were visited (total sample size 6 x 27 = 162). Another approach would be to select the 270 from the six locations (45 from each location). This latter number would seem to be more logical, since one should not be able to draw the same conclusion from less work. However, the two-stage model developed by Leslie (1979) demonstrates that a sample of 477 must be selected from the six branches if a two-stage risk of 5 percent is desired. The samples of 162 and 270 result in risks of 21.9 percent and 10.8 percent respectively. Although empirical evidence is not available, it would be a fairly safe bet that auditors do not increase their sample sizes to the extent illustrated by Leslie in such circumstances and in many cases they would not even consider visiting as many as six out of ten locations (or 60 out of 100, as the case may be).

Given the overly competitive environment facing the profession in all parts of the world at the present time, it is likely that far less work is being done in multilocation situations than in single location situations. In support of this practice it might well be argued that the simple fact that the items are distributed in different locations reduces the risk of error (or of large errors) that would exist if the items were all in one location. Some empirical evidence of this point would certainly be helpful.

The second part of this paper expands the earlier work of Leslie and addresses the issue of evaluating samples when errors have been found. A multilocation example from practice is set out in the third part. The fourth part examines a multilocation application of analytical review (regression analysis) and, in the last part, this is compared with the two-stage sampling approach.

AN ANALYSIS OF THE MULTILOCATION AUDIT PROBLEM

Reiteration of an argument is often more effective than its inherent logic.2

The multilocation audit environment raises two main issues. The first is the extent of work required in terms of the number of locations to be visited and the intensity of work, or sample size, at each location visited. The second issue is results evaluation at the locations visited including an assessment of the effect of those results on the overall audit conclusion. Clearly, these represent the two separate aspects of the general audit sampling problem, that is, the planning and eval-
uation of audit samples. A proper solution to either aspect must also address the other.

Leslie (1979) presents an analysis of the first aspect of the multilocation audit problem that will be used as the basis for the present discussion. Leslie's solution involves the development of a two-stage attribute sampling model. Stage one is the selection of locations and stage two is the drawing of individual items from the selected locations. Leslie's illustration, as noted in the previous section, uses an accounts receivable population with ten locations having decentralized records, but the structure of his approach is easily generalized.

To begin, we shall discuss an audit population with \( D \) dollars in total that is divided equally among \( L \) locations either through its physical location, such as an inventory, or as the result of decentralized accounting records, such as Leslie's accounts receivable example. To achieve his/her sampling objective, the auditor could visit all \( L \) locations and draw a dollar-unit sample, treating the population as if it were at one location. If materiality for the entire audit were \( M \) and the auditor could tolerate a risk \( \beta \) of an incorrect conclusion, then a sample of size \( n^* \), where \( n^* = (-\ln \beta) \times D \div M \), would achieve the desired sampling objective if no errors were detected in the sample (or anticipated throughout the audit).

On what basis could the auditor visit fewer than \( L \) locations? To answer this question, Leslie introduces the location error-size limit (ESL) audit assumption. If the auditor feels that there is other than a very small chance that an error greater than \( M \) could occur in a particular location and not be detected if he/she did not sample that location, then there is no choice but to visit and sample the particular location. However, in practical situations, the chance that a single location, in a multilocation population, will contain a material error often is quite remote. Indeed, for the example presented in the next section, many of the individual location totals are less than half materiality. In any case, without the location ESL assumption, the problem ceases to be one of multilocation sampling.

As one would expect, in practice the location ESL is not an absolute assumption but rather one that should properly be expressed in probabilistic terms. When making a location ESL assumption, the auditor is not saying that there is no chance that if a location's error exceeds the ESL, it will not be detected. Rather, the auditor is saying that, as a result of the inherent risks involved, the assurance available from internal control and the assurance derived from analytical review, there only is a small chance that the auditor will not detect that a location's error exceeds the ESL. This amounts to a Bayesian prob-
ability statement concerning the size of undetected error that may occur at a location. The concept is further illustrated in the Chart, where the ESL is indicated at a point where the Bayesian probability of error greater than the location ESL is quite remote.

The location ESL assumption allows us to introduce a two-stage sampling plan when the ESL < M. The approach for determining the number of locations N to visit and the individual location sample size n is to require that the sampling risk of an audit effectiveness error not exceed \( \beta \). Thus, \( N \) and \( n \) are selected such that

\[
\text{Prob (no errors detected in the sample when error is material)} = \beta.
\]

The formula used to calculate the probability on the left of the foregoing equation will depend upon the location selection method used, the sampling method at selected locations (dollar-unit sampling in this case), and the amount and distribution of error in the population. In the Appendix it is shown that for most reasonable

**Chart. Illustration of Average* Location Error Probabilities**

```
```

Region A has probability .001, say, and Regions A+B have probability .5.

\*Average is taken over the N-dimensional joint probability space of the combined location error probabilities.
location selection methods, the distribution of error that causes the maximum probability to be attained for fixed $N$ and $n$, is that where a portion, $a_i$, of the $L$ locations have error equal to the ESL and the remaining locations are error-free. This result is used implicitly in Leslie (1979).

A conservative choice of $N$ and $n$ is, therefore, one for which the risk of an effectiveness error is calculated assuming $a_i$ locations having error equal to the ESL and the remaining $L-a_i$ not having any error. Accordingly, consideration of the two-stage model developed in Leslie (1979) is justified. In the first stage, the probability of selecting $x$ locations with error equal to the ESL out of the $N$ selected in total is given by a hypergeometric distribution with

$$f(x) = \binom{a_i}{x} \frac{\binom{L-a_i}{N-x}}{\binom{L}{N}},$$

(1)

where $x \leq a_i$ and $f(x) = 0$, $x > a_i$. If $L$ and $a_i$ are large relative to $N$ and $x$, $f(x)$ could be computed using a binomial distribution with parameters $N$ and $p = a_i/L$. This approximation can be useful for large $L$ and $a_i$.

The risk associated with the second stage would be computed conservatively using a Poisson probability distribution. Combining the first and second stages, the function $g(y; L, N, n, a_i)$ where

$$g(y; L, N, n, a_i) = \sum_{x=0}^{\infty} f(x) \exp \left\{ \frac{xESL_nL}{D} \right\} \sum_{j=0}^{\infty} \left( \frac{xESL_nL}{D} \right)^j / j!$$

(2)

gives the probability of detecting $y$ or less errors in a two-stage sample with $N$ selected locations and $n$ items sampled from each location.

Decisions on the extent of sampling can be made by choosing $N$ and $n$ so that

$$g(0; L, N, n, a_i) = \beta.$$  

(3)

As Leslie (1979) has shown, there are two principal results from this model. First, there is a stage one minimum number of locations that must be visited, otherwise the risk, that is, $f(0)$, of having no locations selected that have error is greater than the desired risk $\beta$. Second, for all values of $N < L$, the aggregate item sample size, $nN$, will be greater than $n^*$, the sample size that would have been required had all $L$ locations been visited. In some cases, $nN$ can be several times $n^*$.

Obviously, the specific decisions on $N$ and $n$ depend upon the cost-benefit trade-offs between visiting more locations, which involves set-
up costs, and sampling more extensively at the selected locations. In most practical situations, the set-up costs are significant and, thus, the trade-off favors the lowest possible \( N \).

In large measure, the foregoing analysis has answered the question of the extent of audit work required in a multilocation environment, at least within the constraints of the type of situation considered. It has not, however, dealt with the entire audit sampling problem. Auditors do find errors in their samples and the effects of these detected errors on the audit conclusions must be evaluated. In particular, the actual error point estimate and the upper-error limit must be calculated to reflect the postaudit accuracy of the auditor's knowledge with respect to the population error. Methods for dealing with these issues also will affect the extent of audit work required in a multilocation environment.

Given the multilocation sampling design previously discussed, the following point estimate of error is suggested. Suppose the \( i^{th} \) sample item from the \( j^{th} \) selected location has tainting \( t_{ij} \). We compute our error estimate \( \hat{E} \) as

\[
\hat{E} = \left( \frac{D}{nN} \right) \sum_{j=1}^{N} \sum_{i=1}^{n} t_{ij}.
\]

This estimate also can be written as

\[
\hat{E} = \frac{L}{N} \sum_{j=1}^{N} \frac{D}{nL} \sum_{i=1}^{n} t_{ij},
\]

which is of the form

\[
\hat{E} = \text{(location sampling fraction)} \times \text{(sum of each visited location point estimate of error)}.
\]

The latter formulation is often useful in practical situations. Further, since the locations were selected using simple random sampling and the sample items at the locations were selected using dollar-unit sampling, it is clear that \( \hat{E} \) is an unbiased estimate of the population error.

In common with a number of audit-sampling plans, the dollar-unit sampling strategy requires setting of a basic precision (BP) for sample-planning purposes lower than the materiality set for the audit engagement as a whole. The difference between BP and materiality reflects a planned allowance for detected but uncorrected error, an allowance for the precision increments that occur when errors are detected in attributes-based sampling plans (whether or not errors
are corrected), and a cushion to be slightly conservative. Thus, when errors are detected in this plan, upper-error limits are computed and compared with materiality to make audit judgments concerning the evaluation of the sample results. The appropriate adjustment necessary to arrive at BP from materiality, therefore, will depend upon the methods used for estimating the population error and for computing upper-error limits. Since BP has a direct impact on sample-size calculations, the importance of considering both the planning and evaluation steps in audit sampling becomes clear. Obviously, the two-stage multilocation sampling plan discussed here also requires a calculation of the upper-error limit.

Suppose the auditor has detected γ 100 percent errors in a multilocation sample where \( N \) locations were selected and the sample size was \( n \) per location. The usual approach for calculation of upper-error limits in audit sampling involves computing the maximum value of

\[
E = E_1 + \ldots + E_L,
\]

where \( E_i \) is the error amount at each location, satisfying \( 0 \leq E_i \leq ESL \), and also is constrained to satisfy

\[
\text{Prob (γ or less 100 percent errors)} = \beta.
\]

Note that the probability will be a function of the \( E_i \), the population \( D_i \), the number of locations \( L \), and the sample parameters \( N \) and \( n \). Unfortunately, the rather simple functional properties of the risk function with \( γ = 0 \) do not extend to values of \( γ > 0 \). Thus, the problem of determining the maximum of \( E = E_1 + \ldots + E_L \) cannot automatically be reduced to an enumeration of all boundary cases, that is, those cases where each \( E_i = ESL \) or zero only, since the results of the Appendix will not apply for \( γ > 0 \).

In contrast to the single-location audit sample, for which upper-error limits for samples with 100 percent errors easily can be determined using the binomial or Poisson distributions, even the 100 percent error case is complex in the multilocation situation. If the number of errors detected is large, then classical two-stage calculations similar to those given in Cochran (1977, p. 279) would be suitable. However, for audit applications, the samples are generally too small to provide a sufficient number of errors for the classical methods to be appropriate. The principal difficulty is that the function \( g(γ; L, N, n, a_i) \) may not be appropriate when \( γ > 0 \), since there may be distributions of error in the population that would cause

\[
\text{Prob (γ or less errors)} > g(γ; L, N, n, a_i),
\]
and hence a conclusion drawn using $g$ as the risk constraint would be incorrect.

The actual distribution of error could involve $a_1$ locations with error $\epsilon_1 = ESL$, $a_2$ locations with error $\epsilon_2 < \epsilon_1$, $a_3$ locations error $\epsilon_3 < \epsilon_2$ and so on, and $b$ locations with no error. In this case, the probability of $y$ or less 100 percent errors could be computed using a combined multinomial-Poisson probability formula but the computations would be highly complex for reasonable values of $L$ and $N$. Provided $L$ is large enough, we can approximate the probabilities by replacing the multinomial terms in the probability formula with multinomial terms with parameters $p_i = a_i/L$. Using this approximation, we would have

\[
(7) \quad \text{Prob (} y \text{ or less errors)} = \sum_{k_1=0}^{N-a_1} \cdots \sum_{k_n=0}^{N-a_n} \binom{N}{k_1, \ldots, k_n} (1-p_1-\ldots-p_n)^{N-a_1} \cdots \cdot \cdot \cdot \cdot \cdot \cdot p^n h(y; k_1m_1 + \ldots + k_nm_n),
\]

where $h(y; \theta) = e^{-\theta} \sum_{j=0}^{y} \theta^j/j!$ and $m_n = a_nL/D$ is the expected number of errors in the $n$-sample at each of the locations with error $\epsilon_n$.

The upper-error limit rate would be computed as the maximum value of $R$ where

\[
(8) \quad R = (p_1m_1 + \ldots + p_nm_n)/n, \\
\text{subject to } 0 \leq p_i \leq 1, \ p_1 + \ldots + p_n \leq 1, \text{ and } \ \text{Prob (} y \text{ or less errors)} = \beta.
\]

This formulation of the problem could be solved in particular circumstances with a computerized nonlinear optimization algorithm similar to that used by Fienberg, Neter, and Leitch (1977) in their study of upper-error limit computations for dollar-unit sampling.

Even with this simplification, the problem of determining the maximum for $R$ remains quite formidable. As a practical solution we have used the formulation with $\alpha = 1$, that is, a binomial-Poisson probability formula, in making the upper-error limit calculations for our example in the next section. In this case, $m$ is set equal to $n$ ESL $L/D$ and we solve

\[
(9) \quad \sum_{k=0}^{N} \binom{N}{k} p^k(1-p)^{N-k}e^{-km} \sum_{j=0}^{y} (km)^j/j! = \beta
\]

for $p$ and compute $R = (pm)/n$ as the upper limit on the population error rate. For situations with low values of $y$, this approximation should be reasonably satisfactory.
This upper-error limit calculation provides a solution only when 100 percent errors are encountered. In practice, errors with a variety of taintings will be encountered and, in this case, a modification of the upper-error limit calculations along the lines of the Stringer bound in dollar-unit sampling is suggested. A theoretically more correct model would use multinominal probabilities for both the location categories (the first stage) and the sample taintings (the second stage). Such a model is perhaps far too cumbersome for practical application but it would be useful to explore some of its properties.

In this section, we have outlined the main issues in the multilocation audit-sampling problem including the sample extents and the evaluation of the results of the audit sample. In the next section, these concepts are applied to a practical audit situation.

AN EXAMPLE FROM PRACTICE

A little inaccuracy sometimes saves tons of explanation.

Although it is essential that mathematical models of realistic audit problems be tractable for them to be useful, they must also possess some reasonable correspondence with auditing’s practical realities. This is especially the case for multilocation problems in which a close correspondence of the mathematical model to a realistic situation would render the mathematics intractable. Consequently, a practical illustration of a multilocation audit problem is worthwhile. The following example is drawn from an audit engagement for which one of the authors was the engagement manager.

The client is a specialty retailer dealing in home leisure products. In the audit year, there were 33 stores located in several cities across Canada, with the greatest concentration of stores in Ontario. Since the company’s merchandise is of very small value on an individual item basis, no perpetual records of store inventories were maintained. Inventory at yearend is determined by a third-party inventory count service. As the inventory is quite material, the auditor’s attendance at the stocktaking is an essential step in the audit.

Relevant information on the specialty retailer’s store inventory, together with the key audit planning information, is summarized in Table 1. The store inventory amounts shown are estimates which were based upon the previous year’s amounts. These data were all that were available, for planning purposes, prior to the yearend. The location ESL assumption of $5,000 was considered conservative in this situation given the small dollar value of the inventory items, the
Table 1. Specialty Retailer

Stores Inventory Information

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated yearend store inventories, at cost</td>
<td>$990,000</td>
</tr>
<tr>
<td>Number of stores</td>
<td>33</td>
</tr>
<tr>
<td>Average estimated inventory per store</td>
<td>$30,000</td>
</tr>
<tr>
<td>Largest store inventory — approximately $50,000</td>
<td></td>
</tr>
<tr>
<td>Smallest store inventory — approximately $20,000</td>
<td></td>
</tr>
</tbody>
</table>

Audit Planning Information

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Materiality</td>
<td>$70,000</td>
</tr>
<tr>
<td>Basic precision for sample planning</td>
<td>$52,000</td>
</tr>
<tr>
<td>Tolerable sampling risk for store inventory (i.e., β)</td>
<td>.1</td>
</tr>
<tr>
<td>Location error-size limit assumption</td>
<td>$5,000</td>
</tr>
<tr>
<td>Sample size if one population</td>
<td>44</td>
</tr>
</tbody>
</table>

Table 2. Specialty Retailer

Summary of Multilocation Sample Extents

<table>
<thead>
<tr>
<th>Number of Locations</th>
<th>Sample Size per Location</th>
<th>Total Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>21</td>
<td>126</td>
</tr>
<tr>
<td>7</td>
<td>12</td>
<td>84</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>72</td>
</tr>
<tr>
<td>9</td>
<td>7</td>
<td>63</td>
</tr>
<tr>
<td>10</td>
<td>6</td>
<td>60</td>
</tr>
</tbody>
</table>

The fact that an independent count service was being used, and given the client’s detailed store-by-store inventory and margin analysis. The last two criteria were supplemented by direct communication with the count service and through detailed store-by-store discussions with client management concerning the inventory levels at particular stores.

The planning decisions involving the number of stores to visit and the count extents at each store were made with the aid of the multilocation sample extent procedures discussed in our second section. Given the planning assumptions, it was found that a minimum of six stores would have to be visited. The indicated sample extents for six to ten stores are summarized in Table 2 and they permit an evaluation of cost-benefit trade-offs of visiting more than six stores and decreasing the total test counts. Due to the relatively high set-up cost of visiting a store in the first place, it was decided to attend the inventory count at the minimum number possible.
The selection of the stores to visit was made on a statistical basis using dollar-unit cell-selection from the previous year's list of store inventories. There was no need to bias further the selection toward any particular location as all the stores were fairly close to at least one of the auditor's offices. The result of this selection process was a geographical distribution in the sample that roughly approximated that of the entire population.

The store inventory was taken using the retail method whereby the retail sales price on groups of shelf inventory were added and the total recorded on a ticket. Technically, it would have been possible to select a valid statistical sample using the tickets for each group of stock items. However, the practical limitations of both time and cost constraints made this approach infeasible and the auditor had to settle for a judgmental value-oriented selection procedure.

Obviously the nonrigorous nature of the sample selection process will impair the statistical validity of the conclusions one hopes to draw from the sample. As is a very common occurrence in audit practice, the practicalities of the situation have constrained the approach. Nevertheless, every effort was made to meet the statistical requirements of representativeness that a rigorous approach would provide. Furthermore, there simply is no rational alternative to evaluating the inventory sample's conclusions as if the sample were selected statistically.

After the inventory count was completed and the inventory summarized, a number of issues came to light. The actual store inventory was $984,000 including inventory for a new store (the 34th) which had been stocked but not opened at the company's yearend. The test count results at the stores that were visited are summarized in Table 3. As can be seen from this summarized count information, store 3 had several count problems and the audit staff attending had wisely extended their count procedures to ensure that a large error was not present at that location. Otherwise, the count results do not appear to indicate major problems in the store inventories.

Table 4 summarizes the calculation of the point estimate of error in the store inventory. The net tainting for each store is simply the sum of the reported count taintings and the deemed cell width for each store was computed by dividing the particular store's total inventory at cost by the actual number of items counted at that store (see Table 3). Note that all of the individual store error estimates are well below the location ESL assumption. Two methods are shown for projecting the error detected at the visited stores over the entire store inventory. The second method was used for error summariza-
Table 3. Specialty Retailer

<table>
<thead>
<tr>
<th>Store Number</th>
<th>Store Inventory per Count Service</th>
<th>Count Errors Reported Findings Oversite/(Undersite)</th>
<th>Actual Sample Size Taken</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$30,000</td>
<td>50%, (25%)</td>
<td>21</td>
</tr>
<tr>
<td>2</td>
<td>$57,000</td>
<td>(9%)</td>
<td>21</td>
</tr>
<tr>
<td>3</td>
<td>$27,000</td>
<td>38%, (11%), (31%), (207%)</td>
<td>42</td>
</tr>
<tr>
<td>4</td>
<td>$49,000</td>
<td>(67%)</td>
<td>21</td>
</tr>
<tr>
<td>5</td>
<td>$31,000</td>
<td>10%</td>
<td>21</td>
</tr>
<tr>
<td>6</td>
<td>$41,000</td>
<td>none</td>
<td>21</td>
</tr>
</tbody>
</table>

Table 4. Specialty Retailer

<table>
<thead>
<tr>
<th>Store Number</th>
<th>Net Tainting</th>
<th>Deemed Cell-Width</th>
<th>Store Error Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25%</td>
<td>$1,429</td>
<td>$ 357</td>
</tr>
<tr>
<td>2</td>
<td>(9%)</td>
<td>2,714</td>
<td>(244)</td>
</tr>
<tr>
<td>3</td>
<td>(211%)</td>
<td>643</td>
<td>(1,357)</td>
</tr>
<tr>
<td>4</td>
<td>(67%)</td>
<td>2,333</td>
<td>(1,563)</td>
</tr>
<tr>
<td>5</td>
<td>10%</td>
<td>1,476</td>
<td>148</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$ (2,659)</td>
</tr>
</tbody>
</table>

Projection to total store inventory:
Method 1: (34/6) x (2,659) $15,068
Method 2: ($984,000/235,000) x (2,659) $11,134

As discussed in the second section, the calculation of upper-error limits in the multilocation situation is quite a complex mathematical problem. While a theoretically correct calculation would involve the somewhat sophisticated optimization problem considered earlier, the simpler approach based upon the binomial-Poisson model is more tractable and will be used here. As has been explained, the binomial-Poisson model requires the calculation of the error rate $R = (pm_n)/n$ where $n$ is the location sample size, $m_n$ is the maximum location
expected error based upon the error-size limit assumption and $p$ is the proportion of locations with expected error $m$. The value of $p$ is determined from the formula

$$
(10) \quad \sum_{i=0}^{N} \left( \frac{N!}{i!} \right) p^i(1-p)^{N-i} e^{-m} \sum_{j=0}^{m} (km_j)!/j! = \beta,
$$

where $x$ is the observed number of 100 percent errors.

For the specialty retailer, $N = 6$, $m = 21 \times 5,000 = 30,000 = 3.5$, $\beta = .1$, and $y = 21$. Values of $R = p(3.5)/21$ for various values of $x$ are shown in Table 5. The incremental differences column will be used to compute an upper-error limit using a conservative matching of those incremental differences with error taintings, similar to that used for the Stringer bound in dollar-unit sampling. The resulting calculation for both overstatements and understatements is shown in Table 6. Note that the tainting factors from store 3, where several count errors were detected, have been halved to give effect to the doubling of the sample size at the location. This halving procedure is obviously heuristic in nature but appears to be a reasonable approach in such circumstances. It also was applied indirectly via a lower cell-width in the calculation of the point estimate of error in Table 4.

In computing the upper-error limits, incremental differences in the upper-error-limit rates have been multiplied by the tainting percentages to give the largest value possible. This approach is the same as that employed when computing upper-error limits for dollar-unit samples using the Stringer method.

**Table 5. Summary of Various Population Error Rate Upper Limits**

<table>
<thead>
<tr>
<th>Number of Errors</th>
<th>Upper Limit on Population Error Rate</th>
<th>Incremental Differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.055</td>
<td>N/A</td>
</tr>
<tr>
<td>1</td>
<td>.06</td>
<td>.005</td>
</tr>
<tr>
<td>2</td>
<td>.07</td>
<td>.01</td>
</tr>
<tr>
<td>3</td>
<td>.08</td>
<td>.01</td>
</tr>
<tr>
<td>4</td>
<td>.09</td>
<td>.01</td>
</tr>
<tr>
<td>5</td>
<td>.10</td>
<td>.01</td>
</tr>
<tr>
<td>6</td>
<td>.106</td>
<td>.006</td>
</tr>
</tbody>
</table>

*The upper limits are calculated at a nominal 90 percent confidence level as $(1/21) \times p \times 3.5$, where $p$, the rate of locations with error, is the solution to the equation $\sum_{k=0}^{\infty} \left( \frac{6}{k} \right) p^k(1-p)^{6-k} e^{-3.5m} \sum_{j=0}^{m} (3.5k)!/j! = .1$. 
Table 6. Specialty Retailer

**Calculation of Upper-Error Limits on Store Inventory Sample**

(A) **Overstatements**

<table>
<thead>
<tr>
<th>Error stage</th>
<th>Error rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>0. Basic precision</td>
<td>.055</td>
</tr>
<tr>
<td>1. 50% x .01</td>
<td>.005</td>
</tr>
<tr>
<td>2. 19% x .01</td>
<td>.0019</td>
</tr>
<tr>
<td>3. 10% x .005</td>
<td>.0005</td>
</tr>
</tbody>
</table>

Total UEL error rate = .0624 x $984,000 = $61,402

Gross UEL of overstatements = $61,402

Less: point estimate of understatements* = (15,768)

Net UEL of overstatements at 90% confidence level = $45,634

(B) **Understatements**

<table>
<thead>
<tr>
<th>Error stage</th>
<th>Error rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>0. Basic precision</td>
<td>(.055)</td>
</tr>
<tr>
<td>1. (104%) x .01</td>
<td>(.0104)</td>
</tr>
<tr>
<td>2. (67%) x .01</td>
<td>(.007)</td>
</tr>
<tr>
<td>3. (25%) x .01</td>
<td>(.0025)</td>
</tr>
<tr>
<td>4. (16%) x .01</td>
<td>(.0016)</td>
</tr>
<tr>
<td>5. (9%) x .006</td>
<td>(.0006)</td>
</tr>
<tr>
<td>6. (6%) x .005</td>
<td>(.0005)</td>
</tr>
</tbody>
</table>

Total UEL error rate = (.0779) x $984,000 = $76,063

Gross UEL of understatements = $76,063

Add: point estimate of overstatement = 4,534

Net UEL of understatements at 90% confidence level = $71,429

* Total of negative taintings times the location deemed cell-widths all adjusted for the fact that only 6 locations were visited, i.e., ($984,000/$235,000) x ((25% x 1,429) + (9% x 2,714) + ((11% + 31% + 207%) x 648) + (67% x 2,333)).

As can be seen in Tables 4 and 6, the audit conclusions with respect to the store inventories were consistent with the original expectations. The estimate of error in the stores inventory was about an $11,000 understatement, well below the materiality level for the audit. In addition, the upper-error limits on overstatement and on understatement were within an acceptable range.
A MULTILOCATION APPLICATION OF ANALYTICAL REVIEW

If you torture the data long enough, it will confess.  

The use of analytical review in a multilocation environment requires the same basic audit-planning considerations that are required for alternative audit procedures such as transaction testing and the review and evaluation of internal controls. The overall audit approach for an engagement and its implementation through the planning of specific audit procedures determines the audit objectives to be addressed by analytical review, and it also affects the suitability of analytical review in particular circumstances. These concepts can be illustrated with the specialty retailer example used in the previous section.

In the case of the specialty retailer, store sales are all cash or credit card and hence there is little risk of sales overstatements not being detected by the standard cash audit procedures should such overstatements occur. Additionally, the purchasing was on a centralized basis which allowed the audit of accounts payable to be performed at the head office. On the other hand, the audit of sales for understatements was one area that required attention. Typical approaches for this audit exposure involve a review and evaluation of internal controls together with appropriate compliance testing and some form of analytical review. The internal control analysis would have involved visiting locations.

Possible methods for analytical review include simple comparisons of sales between stores and from one year to the next as well as gross profit ratio analysis. Both of these had been used in the previous year's audit for store sales understatements together with reliance on internal controls with some on-location compliance testing. After examining the alternatives, it was decided that regression analysis would be employed for the sales-understatement audit. At the time, we had been experimenting with the STAR program which we had obtained access to through an agreement with Deloitte, Haskins & Sells.  

Our earlier experiments with the technique indicated that a multilocation environment would be highly suitable for its application.

After some preliminary planning and data collection, we used the software to develop a regression model for store sales as a function of store cost of sales and store operating costs. The model was developed using the annual figures, that is, annual sales, cost of sales, and operating costs, for each store for the previous year. It then was used to predict the annual store sales for each store in the audit year.
As is well known, the STAR software incorporates an audit interface which computes a cut-off threshold for the differences between the regression model predictions and the client's recorded amounts. For differences greater than the cut-off threshold, the software will indicate an unusual situation and a further audit investigation will be required. The cut-off threshold is calculated, based upon an input confidence level and materiality level, explicitly using the prediction interval probabilities from the regression model. The formulation of the STAR investigation rule presented in Kinney and Salamon (1982) can easily be put into the form where an audit investigation is indicated in a particular period if the excess difference $\Delta > 0$, where

$$\Delta = Y_r - Y_b + \max_{\theta < \gamma} (s^* t_{1-\beta,\nu} - M \gamma)$$

and

$Y_r$ is the regression estimate, $Y_b$ the recorded book value, $s^*$ the standard error of prediction for the data point, $\nu$ the degrees of freedom of the regression, $\beta$ the tolerable risk level, and $M$ is the materiality level chosen.

Table 7 contains a reproduction of the audit period report for the sales regression analysis application at the specialty retailer. The report is from a microcomputer regression analysis package developed at Clarkson Gordon for use in audit engagements. This package, known as CGDATA (an acronym for Clarkson Gordon Data Analysis Technique for Auditors), is similar in operation to STAR and presently employs the same investigation rule. For the example presented, basic precision of $52,000$ and a Poisson confidence factor of $2$ were used in the audit interface calculations.

In the audit, the excess differences at observations 45 and 65 were followed up with inquiries of senior client management as to the cause of the unusually low sales values. The store represented in observation 45 was to be closed in the near future and was used as a clearance center during the year. The store represented in observation 65 was a fairly new one that had been charged for the cost of inventory that was shipped to the company's head office warehouse. This error involved an internal allocation that did not affect the financial statements.

The quality of the regression-based analytical review was felt to be considerably superior to that used in the previous year and therefore, it was felt appropriate to reduce the extent of reliance on internal controls and to omit the compliance procedures. The cost of applying the regression approach was roughly the same as what had been
Table 7. Statistical Information for Selected Equation Dependent Variable: Store Sales

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Coefficient</th>
<th>T-Statistic</th>
<th>R(l) Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-746.20925</td>
<td>-1.472</td>
<td></td>
</tr>
<tr>
<td>Cost of Store S</td>
<td>1.83013</td>
<td>63.389</td>
<td>0.0842</td>
</tr>
<tr>
<td>Operating Costs</td>
<td>-0.88091</td>
<td>-2.734</td>
<td>0.0842</td>
</tr>
</tbody>
</table>

R-Squared: 0.9933

Standard Error: 655.1426

Diagnostics Analysis for Selected Equation

<table>
<thead>
<tr>
<th>Obs. Num.</th>
<th>Observed Value</th>
<th>Estimated Value</th>
<th>Residual Value</th>
<th>Studentized Residual</th>
<th>WSSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>386.13</td>
<td>379.77</td>
<td>6.36</td>
<td>1.0</td>
<td>368</td>
</tr>
<tr>
<td>2</td>
<td>191.14</td>
<td>199.19</td>
<td>-8.05</td>
<td>-1.3</td>
<td>75</td>
</tr>
<tr>
<td>3</td>
<td>205.56</td>
<td>205.56</td>
<td>-0.00</td>
<td>0.0</td>
<td>57</td>
</tr>
<tr>
<td>4</td>
<td>265.10</td>
<td>257.83</td>
<td>7.27</td>
<td>1.1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>254.17</td>
<td>250.82</td>
<td>3.35</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>262.55</td>
<td>264.72</td>
<td>-2.17</td>
<td>-0.3</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>242.88</td>
<td>249.33</td>
<td>-6.45</td>
<td>-1.0</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>173.34</td>
<td>182.50</td>
<td>-9.16</td>
<td>-1.4</td>
<td>121</td>
</tr>
<tr>
<td>9</td>
<td>379.21</td>
<td>372.74</td>
<td>6.47</td>
<td>1.0</td>
<td>321</td>
</tr>
<tr>
<td>10</td>
<td>284.06</td>
<td>281.14</td>
<td>2.92</td>
<td>0.5</td>
<td>15</td>
</tr>
<tr>
<td>11</td>
<td>304.49</td>
<td>306.24</td>
<td>-1.75</td>
<td>-0.3</td>
<td>59</td>
</tr>
<tr>
<td>12</td>
<td>427.52</td>
<td>435.10</td>
<td>-7.58</td>
<td>-1.3</td>
<td>759</td>
</tr>
<tr>
<td>13</td>
<td>454.63</td>
<td>454.21</td>
<td>0.42</td>
<td>0.1</td>
<td>979</td>
</tr>
<tr>
<td>14</td>
<td>202.48</td>
<td>190.11</td>
<td>12.37</td>
<td>1.9</td>
<td>95</td>
</tr>
<tr>
<td>15</td>
<td>220.02</td>
<td>212.11</td>
<td>7.91</td>
<td>1.2</td>
<td>41</td>
</tr>
<tr>
<td>16</td>
<td>215.68</td>
<td>213.63</td>
<td>2.05</td>
<td>0.4</td>
<td>40</td>
</tr>
<tr>
<td>17</td>
<td>154.33</td>
<td>155.63</td>
<td>-1.32</td>
<td>-0.2</td>
<td>231</td>
</tr>
<tr>
<td>18</td>
<td>334.24</td>
<td>341.93</td>
<td>-7.69</td>
<td>-1.2</td>
<td>175</td>
</tr>
<tr>
<td>19</td>
<td>305.66</td>
<td>300.22</td>
<td>5.44</td>
<td>0.8</td>
<td>51</td>
</tr>
<tr>
<td>20</td>
<td>239.67</td>
<td>238.26</td>
<td>1.41</td>
<td>0.2</td>
<td>6</td>
</tr>
<tr>
<td>21</td>
<td>337.60</td>
<td>343.40</td>
<td>-5.80</td>
<td>-0.9</td>
<td>189</td>
</tr>
<tr>
<td>22</td>
<td>160.39</td>
<td>158.05</td>
<td>2.34</td>
<td>0.4</td>
<td>226</td>
</tr>
<tr>
<td>23</td>
<td>271.66</td>
<td>261.41</td>
<td>10.25</td>
<td>1.6</td>
<td>1</td>
</tr>
<tr>
<td>24</td>
<td>203.38</td>
<td>198.91</td>
<td>4.47</td>
<td>0.7</td>
<td>75</td>
</tr>
<tr>
<td>25</td>
<td>194.02</td>
<td>210.12</td>
<td>-16.10</td>
<td>-2.5</td>
<td>49</td>
</tr>
<tr>
<td>26</td>
<td>241.33</td>
<td>238.64</td>
<td>2.69</td>
<td>0.4</td>
<td>3</td>
</tr>
<tr>
<td>27</td>
<td>151.47</td>
<td>156.67</td>
<td>-5.20</td>
<td>-0.8</td>
<td>206</td>
</tr>
<tr>
<td>28</td>
<td>198.24</td>
<td>194.44</td>
<td>3.80</td>
<td>0.6</td>
<td>67</td>
</tr>
<tr>
<td>29</td>
<td>233.84</td>
<td>243.00</td>
<td>-9.16</td>
<td>-1.5</td>
<td>1</td>
</tr>
<tr>
<td>30</td>
<td>225.95</td>
<td>224.36</td>
<td>1.59</td>
<td>0.3</td>
<td>12</td>
</tr>
<tr>
<td>31</td>
<td>193.35</td>
<td>194.02</td>
<td>-0.67</td>
<td>-0.1</td>
<td>67</td>
</tr>
<tr>
<td>32</td>
<td>179.98</td>
<td>180.23</td>
<td>-0.25</td>
<td>0.0</td>
<td>103</td>
</tr>
</tbody>
</table>
Table 7 (continued)

Other Statistics

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Observed Value</th>
<th>Critical Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skewness</td>
<td>-0.798</td>
<td>1.807</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>-0.042</td>
<td>1.636</td>
</tr>
<tr>
<td>Kolmogorov-Smirnov DN</td>
<td>0.092</td>
<td>0.240</td>
</tr>
</tbody>
</table>

Audit Period Report

<table>
<thead>
<tr>
<th>Obs. Num.</th>
<th>Recorded Value</th>
<th>Predicted Value</th>
<th>Difference</th>
<th>Excess Difference</th>
<th>Optional Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>33</td>
<td>$39915</td>
<td>$40650</td>
<td>$-735</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>18138</td>
<td>18186</td>
<td>-48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>21645</td>
<td>21995</td>
<td>-350</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>29762</td>
<td>29442</td>
<td>320</td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>28902</td>
<td>28852</td>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>28662</td>
<td>28693</td>
<td>-31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>25186</td>
<td>25310</td>
<td>-124</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>17965</td>
<td>18093</td>
<td>-128</td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>41647</td>
<td>40573</td>
<td>1074</td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>30010</td>
<td>30715</td>
<td>-705</td>
<td></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>33910</td>
<td>34010</td>
<td>-100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>48559</td>
<td>48900</td>
<td>-341</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>42084</td>
<td>43770</td>
<td>-1686</td>
<td>$482</td>
<td>12</td>
</tr>
<tr>
<td>46</td>
<td>21667</td>
<td>22060</td>
<td>-393</td>
<td></td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>25794</td>
<td>25194</td>
<td>600</td>
<td></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>23637</td>
<td>23776</td>
<td>-139</td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>16848</td>
<td>16728</td>
<td>120</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>37500</td>
<td>37946</td>
<td>-446</td>
<td></td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>32933</td>
<td>33274</td>
<td>-341</td>
<td></td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>24789</td>
<td>25495</td>
<td>-706</td>
<td></td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>36616</td>
<td>36410</td>
<td>206</td>
<td></td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>16889</td>
<td>15931</td>
<td>458</td>
<td></td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>29647</td>
<td>29954</td>
<td>-307</td>
<td></td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>21785</td>
<td>21380</td>
<td>405</td>
<td></td>
<td></td>
</tr>
<tr>
<td>57</td>
<td>20426</td>
<td>20176</td>
<td>250</td>
<td></td>
<td></td>
</tr>
<tr>
<td>58</td>
<td>26839</td>
<td>26801</td>
<td>38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>59</td>
<td>19432</td>
<td>19831</td>
<td>-399</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>24260</td>
<td>24606</td>
<td>-346</td>
<td></td>
<td></td>
</tr>
<tr>
<td>61</td>
<td>31732</td>
<td>32273</td>
<td>-541</td>
<td></td>
<td></td>
</tr>
<tr>
<td>62</td>
<td>28918</td>
<td>29243</td>
<td>-325</td>
<td></td>
<td></td>
</tr>
<tr>
<td>63</td>
<td>23695</td>
<td>23404</td>
<td>291</td>
<td></td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>25102</td>
<td>25325</td>
<td>-223</td>
<td></td>
<td></td>
</tr>
<tr>
<td>65</td>
<td>18499</td>
<td>21874</td>
<td>-3375</td>
<td>2236</td>
<td>8</td>
</tr>
</tbody>
</table>

| Total     | $912693        | $920868         | $-8175     | $2718             | 20                   |
incurred in the previous year solely on the store gross profit ratio analysis.

THE NATURE OF ANALYTICAL REVIEW

The example presented in the previous two sections gives us an opportunity to compare and contrast multilocation audit sampling with analytical review. One's first reaction might be to conclude that the only similarity between the audit procedures is the multilocation environment. Our view, however, is that the two procedures are similar at a more fundamental structural level.

The multilocation audit-sampling procedure was a two-stage sampling approach in which the auditor makes a selection of locations and then samples within the selected locations. The location selections and samples are made at random under the auditor's control. The sample results are projected to give a point estimate of the error in the multilocation population and the accuracy of that estimate is evaluated using an upper-error-limit calculation.

"Analytical review procedures are substantive tests of financial information made by a study and comparison of relationships among data" (AICPA 1984, Section 318.02). Analytical review requires the auditor either explicitly or implicitly to make a prediction of a certain financial statement result or, perhaps, a derived statistic such as a ratio, and compare that prediction with the observed value of the corresponding information derived from the audit period financial statement information. Significant differences indicate areas where further audit investigation is required. Thus, significance is defined in terms of the auditor's materiality and required level of audit assurance. The regression analysis approach illustrated in the previous section is an analytical review procedure in which the prediction is made explicitly. If the auditor were to compare, say, a previous year's financial statement item with its value in the audit year, then his/her implicit prediction in the audit year is that previous year's result. Similar comments apply to most forms of ratio analysis.

The analytical review process provides the auditor with a tool which allows him/her to detect errors in financial information through an analysis of their effects on summarized accumulations of individual transactions. These summarized accumulations contain random influences from a variety of factors. In virtually all realistic applications of analytical review procedures, the expectation is that the actual results will not agree exactly with the prediction but will differ due to these random influences. For example, individual stores in the
specialty retailer will differ from each other in the mix of the merchandise they sell. Accordingly, since the margins for the merchandise vary within a narrow range, one would not expect all stores to have exactly the same gross profit percentages. Rather, we would expect to see a variation in the gross profit percentages from one store to the other, albeit in a fairly narrow range.

Conceptually, the prediction aspect of the analytical review process can be viewed as that of developing a model in which the statistic being predicted is expressed as the sum of specified influences and random influences, that is,

\[
\text{Analytical review statistic} = \text{effects of specified influences} + \text{effects of other random influences.}
\]

In the foregoing conceptual equation, the analytical review statistic could be a sales prediction from a regression equation or a gross profit ratio. The "effects of specified influences" would represent the independent variables (that is, the equation, in the case of regression) which, in the case of gross profit-ratio analysis, could be sales and cost of sales. The random influences simply represent all other factors not explicitly recognized as specific influences. This corresponds to the theoretical stochastic term in a regression model.

To illustrate the effect of recognizing the random influence aspects of analytical review, consider the previous section's store sales example in which a prediction of a particular store's annual sales is being compared with the recorded amount. Let us suppose that the recorded amount for the store contains a significant error. Under a nonstatistical viewpoint, this error would cause a significant fluctuation of the recorded amount from the prediction and would therefore be detected. However, there are random influences in the relationship used for analytical review. Thus, the recorded amount for that month also may be affected by unusual transactions which are correct but obscure the effect of the error. Depending upon the magnitude of these unusual transactions, detection of the error may not be certain.

The point of this discussion is that most, if not all, analytical review procedures introduce a risk of failing to identify those components (that is, locations, months, and so on) which contain error, even if that error is significant. At least in theory, this risk should be combined with the risk of failing to detect the specific error once a component has been identified as unusual by the analytical review. This latter risk arises from the nonsampling risk of accepting an incorrect
explanation of an unusual fluctuation and, in some cases, the sampling risk associated with a detailed sample of transactions.

Analytical review can, therefore, be viewed as a two-stage audit procedure in which the first stage identifies those components likely to contain error and the second stage is an investigation of the identified components intended to detect any error should it actually exist. In theory, each of these stages introduces a risk element into the process that must be combined to determine the level of audit assurance that can be obtained from analytical review.

The analogy between the multilocation audit-sampling problem and the analytical review process should now be clear. Both audit procedures are two-stage sampling procedures. The essential difference between the two is that the first-stage selection in analytical review depends upon a probability mechanism that is not under the auditor’s control. In the case of analytical review, the auditor cannot apply a randomized approach to select the components for which further investigation is required, as is the case for a multilocation audit-sampling design. Instead, he/she employs the stochastic element of the relationship which is specified by the auditor for the analytical review. In this case, the probability of selection for a particular component will depend upon the model specified, the data used to develop it, and the amount of error in the component. Despite this important difference between the two audit procedures, they remain structurally quite similar.

This structural similarity suggests some rather obvious implications for analytical review. Regardless of the nature of the second-stage procedure, which could be either sampling or inquiry-based investigation, the sampling risk associated with the first stage makes it necessary to project errors detected in a particular component during the second stage to the other components in the population. In fact, if the second stage were certain to detect the error in a component should that error exist, there still would be an error projection due to the first-stage risk.

In addition to projecting the error detected in a component over the other components in the population, the analogy between multilocation audit sampling and analytical review also requires that an assessment of the sufficiency of the analytical review be made through computation of upper-error limits at the tolerable level of audit risk. If analytical review provides any audit assurance at all, then it must, in effect, provide an upper limit on possible error at an acceptable level of risk of being incorrect. When errors are detected with the aid of analytical review, then the upper-error limit, before any client
corrections, must increase. Alternatively, the auditor can maintain the upper-error limit at the no-error level but the risk of being incorrect will necessarily increase.\textsuperscript{10}

For most conventional analytical review procedures, such as comparison and ratio analysis, the projection of detected errors and the calculation of upper-error limits is highly problematic and one would doubt that it is done in practice that often, if at all. However, the conclusions just arrived at indicate quite clearly that, whenever there is a risk that the first stage of the analytical review procedure will fail to require an investigation of a component when error exists, then, in principle, such projections and upper-limit calculations must be made. In many circumstances, the auditor can, however, employ statistical methods to model the relationship that is used in the analytical review. Although the underlying probability structure for the first-stage selection process is not controlled by the auditor, if the auditor is able to develop a reasonable model for that structure, then he/she will be able to make explicit risk calculations.

The regression analysis example presented in the previous section illustrates an analytical review methodology in which a complete description of the modeling process must include the model for the random influences as well as that of the specific predictor variables in the relationship. In this case, there is a theoretical basis for calculating the risk of not investigating a particular component when that component contains a given amount of error, perhaps zero error. Thus, in principle, projections of error and upper-error-limit calculations could be made.

The nature of error projections and upper-error-limit calculations in regression-analysis-based analytical review essentially remains an unsolved research problem. However, one approach to projecting errors which we feel may have some promise is the following.\textsuperscript{11} If we are prepared to assume that the second-stage risk in an analytical review is zero, that is, when we investigate a particular component we are certain to detect the error in that component, then we may calculate the probability of detecting that component error using the prediction interval probabilities from the regression analysis. For example, suppose that we investigate a particular component and detect understatement error $E > 0$. A projection of error from that component would be made with the formula,

$$E = \text{Prob (investigation when error is } E).$$
corrections, must increase. Alternatively, the auditor can maintain
the upper-error limit at the no-error level but the risk of being
incorrect will necessarily increase.\textsuperscript{10}

For most conventional analytical review procedures, such as com-
parison and ratio analysis, the projection of detected errors and the
calculation of upper-error limits is highly problematic and one would
doubt that it is done in practice that often, if at all. However, the
conclusions just arrived at indicate quite clearly that, whenever there
is a risk that the first stage of the analytical review procedure will
fail to require an investigation in a component when error exists,
then, in principle, such projections and upper-limit calculations must
be made. In many circumstances, the auditor can, however, employ
statistical methods to model the relationship that is used in the
analytical review. Although the underlying probability structure for
the first-stage selection process is not controlled by the auditor, if the
auditor is able to develop a reasonable model for that structure, then
he/she will be able to make explicit risk calculations.

The regression analysis example presented in the previous section
illustrates an analytical review methodology in which a complete
description of the modeling process must include the model for the
random influences as well as that of the specific predictor variables
in the relationship. In this case, there is a theoretical basis for
calculating the risk of not investigating a particular component when
that component contains a given amount of error, perhaps zero error.
Thus, in principle, projections of error and upper-error-limit calcula-
tions could be made.

The nature of error projections and upper-error-limit calculations
in regression-analysis-based analytical review essentially remains an
unsolved research problem. However, one approach to projecting
errors which we feel may have some promise is the following.\textsuperscript{11} If
we are prepared to assume that the second-stage risk in an analytical
review is zero, that is, when we investigate a particular component
we are certain to detect the error in that component, then we may
calculate the probability of detecting that component error using the
prediction interval probabilities from the regression analysis. For
example, suppose that we investigate a particular component and
detect understatement error \( E > 0 \). A projection of error from that
component would be made with the formula,

\begin{equation}
E = \text{Prob (investigation when error is } E)\).
\end{equation}
Using the notation introduced in the previous section, the foregoing probability can be calculated as \( \gamma \) where

\[
\gamma = \frac{(\Delta + Y_\alpha - Y + E)}{s^*}.
\]

An overall projection of error could be made by taking the total of the individual component error projections obtained as just described.

Despite the restrictive conditions concerning the second-stage risk, the foregoing error projection procedure has strong intuitive appeal. Under the specified conditions, one would expect the estimation procedure to be unbiased in the sense that, on average, the estimated error would equal the actual population error. Obviously the entire area requires considerable study.

Finally, since the discussion in this section has suggested a two-stage sampling model for analytical review, which can be formalized through the use of statistical methods, a natural possibility to consider is the use of regression analysis to select the locations to visit in a multilocation audit environment. Such an approach fits within the framework developed in the second section, as an improvement in the quality of the analytical review procedure that formed part of the basis for the location error-size-limit assumption in the first place. This quality improvement should allow the auditor to reduce the ESL below the amount that he/she would otherwise have to accept and, in some cases, may eliminate the need for sampling under the multilocation audit model entirely. For example, if, as a result of applying regression analysis, the auditor concludes that the ESL is so low that a material error in the population would be highly improbable, then there would be no need to visit locations other than those identified by the analytical review. In some cases, however, the reduction in the ESL may not be sufficient to allow for such a large reduction in the number of locations to be visited and a selection of locations independent of the analytical review procedure would be required.

The combined type of multilocation audit approach suggested by this analysis does not appear to occur in practice that often when dealing with audit procedures directed at balance sheet items. The principal reason for this is one of timing and the practical nature of audit procedures. The specialty retailer example illustrates this quite clearly. The use of regression analysis to lower significantly the ESL for the inventory count is not very practical since inventory is determined largely by the client's count and, thus, the information necessary to complete the analytical review would not be available until a week or so after the count date. Thus, it is not possible to
make use of the analytical review results to select locations for count attendance.

CONCLUSION

One of the major advantages of being practitioners presenting a paper at an academic symposium is that we really do not have to solve any problems — we need only create them. Multolocation audit sampling is fraught with difficulties both from a technical, mathematical perspective and from the viewpoint of the practitioner in the field. In this paper, we have endeavored to outline what we perceive to be the major problems and we have suggested a number of approaches to their solution.

Our analysis of the multolocation audit-sampling problem provides a reasonably satisfactory method for determining sample size and for calculating point estimates of the population errors. The method we have illustrated for calculating the upper-error limit is perhaps useful provided care is taken to ensure that, in any application, the location ESL assumption is not violated. However, further research is clearly required to develop fully an appropriate method for such computations. Since the calculation method for upper-error limits will affect the determination of sampling basic precision, we feel this is an important area for research.

Our comparison of analytical review procedures with the multolocation audit-sampling environment, or perhaps more generally, multicomponent audit sampling, has suggested the use of a common two-stage sampling framework. In this framework, the identification of components with significant fluctuations introduces a risk of failing to detect components with error. Although this risk is not a typical sampling risk arising from randomization and, therefore, cannot be controlled by the auditor, it remains a real audit risk that must be considered. The use of statistical methods such as regression analysis allows the auditor to model these first-stage risks and, in principle, to incorporate the effect of these risks in the evaluation of the results of analytical review. Research on statistical methods for analytical review should recognize this two-stage sampling framework and address the key issues of calculating point estimates of error and determining upper-error limits.

APPENDIX. AN ANALYSIS OF THE ZERO ERROR CASE IN MULTLOCATION AUDIT SAMPLING

Let us consider a multolocation population with \( L \) locations and such that the error in location \( i \) is \( E_i \). Suppose we wish to draw a two-stage sample
from this population with the first stage being a selection of $N$ of the locations and in the second stage, a dollar-unit sample of size $n$ from each of the selected locations. We assume that the Poisson approximation is reasonable for probability calculations for the second-stage sample.

If we denote the probability of selecting locations $i_1, \ldots, i_6$ by $\pi(i_1, \ldots, i_6)$, then the probability that our two-stage sample does not detect any sample items in error is given by

$$\text{Prob (0 errors)} = \sum_{\pi \in \pi_6} \frac{\pi(i_1, \ldots, i_6)}{\exp(-n(E_{i_1}/D_{i_1} + \ldots + E_{i_6}/D_{i_6}))},$$

where $D_i$ is the dollar amount of the population at location $i$. We require $0 \leq E_i \leq ESL_i$, where $ESL_i$ is a constant location error-size limit assumption that may differ between locations.

Suppose that $E = \sum_{i=1}^{L} E_i$, where $E$ is some fixed amount. If we consider the right-hand side of Equation (A.1) as a function of the $E_i$, we may properly consider the problem of determining the maximum value of the function $f$, where

$$f(E_1, \ldots, E_L) = \sum_{\pi \in \pi_L} \frac{\pi(i_1, \ldots, i_L)}{\exp(-n(E_{i_1}/D_{i_1} + \ldots + E_{i_L}/D_{i_L}))},$$

subject to

$$E = \sum_{i=1}^{L} E_i$$

and

$$0 \leq E_i \leq ESL_i, \quad i \in \{i_1, \ldots, L\}.$$  

Equation (A.3) can be used to substitute for $E_i$ in the formula (A.2) for $f$. This optimization problem can be shown not to have any relative maximum in the interior of the domain for the $E_i, i \in \{2, \ldots, L\}$ as follows. From Equation (A.2), noting (A.3), it follows that the diagonal elements $h_i$ of the hessian of the foregoing optimization problem are all of the form

$$h_i = \frac{\partial^2 f}{\partial E_i^2} = \sum_{\alpha} a_{\alpha} \exp(-n \sum_{A \neq i} E_A/D_A),$$

where all $a_{\alpha} > 0$ and the $A(\alpha)$ represent appropriate indexing sets for each $\alpha$. This implies that $h_i > 0$ for all $i$ and hence the trace of the hessian must be nonnegative. Thus, at all points of the domain at least one of the eigenvalues of the hessian is positive. At any interior critical point of $f$, therefore, the hessian cannot be negative definite. Moreover, there will be at least one principal direction along with the objective function $f$ which can be shown to increase from its value at the critical point (that is, that corresponding to the positive eigenvalue). This proves the claim that the solution to the maximization problem occurs at a boundary point of the admissible domain for the $E$, and repetition of the foregoing argument
allows us to conclude that we must have all the \( E \) equal to \( ESL_{a_i} \), equal to zero, or equal to \( E - ESL_{a_i} \) for some \( i_1, \ldots, i_k \).

Consequently, when planning a multilocation audit sample, the risk calculations for Equation (A.1) need only be made for a finite number of possible error distributions in the population. With \( E \) equal to our sampling basic precision together with a particular choice of location selection method, which determines the \( \pi(i_1, \ldots, i_k) \), and of parameters \( N \) and \( n \), only a finite number of values of Prob (0 errors) need be computed. The maximum of these values can be compared with the tolerable sampling risk to ensure that the planned approach meets the audit requirements.

In the case considered in the second section, the \( ESL_i = ESL \) and \( D_i = D/L \) for all \( i \). Furthermore, the selection of locations is assumed to be based upon simple random sampling without replacement. Under these conditions, it is clear that the boundary case solution to the corresponding optimization problem will lead to \( a_i \) locations with error equal to \( ESL_{a_i} \), where \( a_i \) is the first integer greater than \( a_i \) locations with sampling basic precision divided by \( ESL \), and the remaining locations are error-free. This is exactly the basis for Leslie's model (Leslie 1979).

NOTES

1. These quotations have been taken from the pamphlet *Quotes, Damned Quotes, and . . .*, compiled by John Bibby.

2. See Note 1.

3. The assumption that all locations have equal size is artificial and rarely occurs in practice. However, as an approximation it can be quite useful provided locations are partitioned into groups of roughly the same exposure to error.

4. As \( y \) increases, one must eventually question the validity of the error-size-limit assumption which effectively limits the maximum error to \( N \times ESL \). The entire analysis becomes invalid if the error-size-limit assumption is rejected.

5. The similarity is in the conservative nature of the error-ranking process used in the suggested evaluation approach discussed in the third section.

6. Attributed to Saki, see Note 1.

7. The nonhomogeneity of the locations was therefore addressed through the location selection method. This affects the upper-error-limit calculations which will necessarily be based upon an approximate theoretical model.

8. Attributed to Ronald Coase, see Note 1.

9. We wish to thank Deloitte, Haskins & Sells for giving us the opportunity to use their software, for their cooperation in providing training in its use, and for technical explanations of its operations. This opportunity allowed us to make a quantum leap in our research into the viability of regression analysis in our audit practice.

10. This is essentially the conclusion obtained empirically in Kinney and Salamon (1982) with respect to the STAR investigation rule in a regression analysis analytical review. The generalization to practically all forms of analytical review should be clear.

11. A similar error projection approach was suggested by Gary Holstrum who was discussant of the Kinney and Salamon (1982) paper when it was
presented at the Conference on Audit Risk held in May 1982 at the University of North Carolina at Chapel Hill.

REFERENCES


Models for Multilocation Audits: Discussion

The Aldersley-Leslie (A-L) paper provides an interesting discussion of the problem of statistically sampling from populations which are dispersed over several locations. The authors make three useful contributions to the scholarly literature. First, they identify the potential difficulties in controlling, statistically, overall inferential risks when the selection of locations to be sampled (first-stage sampling) is not made statistically. Second, they propose a two-stage sampling model which enables the auditor to control sampling risks associated with substantive testing in a multilocation environment. Finally, the explicit recognition of the fixed-cost component of sampling (which extant studies largely ignore) is a contribution to modeling.

Like other sampling activities, multilocation sampling requires the auditor to make (implicit or explicit) risk-return trade-offs. In the multilocation context, the presence of fixed start-up costs can make statistical sampling at all locations too costly and also may delay completion of the audit. When auditors visit only a subset of locations, however, a problem arises: the auditor has to make inferences about the entire set of location populations based upon statistical samples which are obtained from a subset of locations. A-L argue that, when
the location site selection process is made judgmentally (as apparently is done in practice), the auditor faces essentially the same problem as when judgmental sampling methods are employed — inferential risks cannot be quantified and thereby controlled. To overcome this problem, a two-stage sampling model is presented by A-L in which the locations are selected on a statistical basis. In the following sections, I discuss the four issues in relation to their model (enumerated below):

(1) Sampling cost structure in multilocation environments;
(2) Coordination of compliance and substantive testing;
(3) Statistical control of efficiency risks; and
(4) Basis for statistically selecting locations.

Subsequently, I describe potential extensions to the A-L model and provide concluding remarks.

**SAMPLING COSTS**

Two-stage sampling plans originally were developed by statisticians to reduce the fixed-cost component of sampling, such as travel costs, by explicitly recognizing such natural groupings as households and towns in conducting large-scale surveys (see Cochran 1977). The basic idea in an auditing context is to achieve overall sampling risk objectives by statistically identifying a small number of locations which can be sampled very intensively. Such sampling plans effectively trade off the reduction in fixed costs against increased variable costs. An efficient trade-off between fixed and variable sampling costs, however, requires an explicit consideration of both types of sampling costs. Furthermore, recognition that the fixed and variable costs of sampling will not necessarily be the same at each branch location also would seem to be important.

The present version of the A-L model assumes implicitly that (1) variable sampling costs are negligible vis-à-vis fixed sampling costs at all locations and (2) that fixed sampling costs are the same for each branch location. Both assumptions potentially limit the applicability of the present version of their model.

**EFFICIENCY RISK**

The A-L model enables auditors to control statistically the probability of making effectiveness errors (that is, concluding that the account balance is fairly presented when it actually is intolerably misstated). Although effectiveness risk is the auditor’s primary con-
cern, the problem of incorrectly concluding that a fairly presented account balance is misstated is also of importance. In such circumstances, Elliott and Rogers (1972) have argued that unnecessary audit procedures will be performed. Since the auditor will be forced to absorb most if not all of these costs in a competitive environment, control over efficiency risk is essential. Hence a useful extension of the A-L model would be to control efficiency and effectiveness risks concurrently.

COMPLIANCE AND SUBSTANTIVE TESTS

The present version of the A-L model focuses on the auditor’s substantive testing problem. Implicit in the model’s application is the assumption that compliance tests are performed at the appropriate branch locations prior to substantive testing. In fact, A-L argue that the auditor’s compliance test information will provide the basis for excluding certain branches from the first-stage location sample selection process. My question is the following: If the fixed costs of visiting the branch locations are really much more important than the variable sampling costs (as A-L assume), why not perform substantive tests at a given branch immediately after compliance tests are performed? It would seem that, when possible, coordinating the timing of compliance and substantive tests would be the most cost-effective testing policy.

My point in raising the compliance-testing issue is not to criticize the A-L paper, but instead to identify one possible avenue for an extension of their model. Along a slightly different vein, I believe that the same basic sampling cost issues are relevant to compliance testing at multilocations, and that a modified version of the A-L model could be applied to multilocation compliance tests.

LOCATION SELECTION

The statistical selection of branch locations which underlies the A-L two-stage sampling model is appropriate; indeed such selection is essential to the quantification and control of sampling risks. Overall sampling-risk quantification in the A-L model is based upon the simplifying assumption that each location (included in the first-stage sample space) is to be selected with equal probability and that the total number of branches to be visited is fixed a priori. If all branches are expected to be homogeneous with respect to business transactions, size, personnel, management, and internal control environments, and so on, such an assumption obviously is quite defensible. However, to the extent that there are interbranch similarities, a sequential plan
would provide further efficiency gains. Specifically, a sequential plan
would allow the auditor to utilize sample information obtained from
one location in deciding whether to visit additional branches, rather
than committing ex ante to visit a predetermined number of locations
as the A-L plan requires.

Alternatively, if the branches are expected to have heterogeneous
features, then recognition of such differences would be of utmost
importance in selecting the locations and also in evaluating ex post
sampling risks. The authors obviously are sensitive to this point and
suggest that, when feasible, analytical review can help the auditor to
select the branches to be sampled. However, the linkage between
analytical review and location selection has not yet been formalized
in their model. My suggestion to the authors is to formalize this
linkage by adopting a Bayesian decision-theory framework. A Bayesian
framework would permit the auditor to incorporate analytical review
and other presubstantive testing information in the form of prior
probability distributions.

EXTENSIONS

There are several potential advantages associated with a Bayesian
approach to the multilocation audit problem. If one looks carefully
at the essential features of such a sampling environment, however,
many of the key issues are similar to the problem of allocating audit
effort among different accounts addressed by Scott (1973). In fact,
the multilocation allocation problem can be viewed as being a special
case of the latter. First, due to greater similarity in the underlying
transactions and internal controls, the sampling units constituting the
various locations are likely to be much more homogeneous than in
the multiaccount context. Second, the auditor's reporting options are
more constrained than in the multiaccount context. For example, in
reporting on published financial statements, the auditor not only
attests to the fair presentation of the financial statements taken as a
whole, but implicitly attests to the fairness of the individual account
balances constituting the statements. If appropriate, the auditor has
the ability to qualify an opinion with respect to a particular account
balance. Such a reporting option, however, ordinarily would not be
available with respect to a location. Hence, the auditor's loss function
in a Bayesian multilocation model would be a special case of the loss
function in Scott's model in that the loss function would be defined
with respect to the aggregate monetary error rather than separately
for each branch. Third, a multiprocess (population) model such as
the one developed by Scott (1973) has the generality to handle heterogeneous branches and subsidiaries (one of the applications proposed by A-L).

CONCLUSIONS

In summary, A-L should be commended for identifying the potential problems which can arise in applying statistical sampling plans in multilocation audits and for proposing a two-stage sampling plan which allows auditors to obtain better control over audit effectiveness errors. I hope that the A-L framework will provide the starting point for other refinements which will further enhance the model's potential applicability and existing contributions.

REFERENCES


Models for Multilocation Audits: Discussion

I am pleased to have been asked to comment on the Aldersley-Leslie paper because it addresses an important issue that previously has been ignored in the literature. At least the discussion has been started so perhaps the profession (both practitioners and academics) will begin to think about and address the issue of work allocation across locations. There is, however, a general impression that this paper might leave that concerns me.

Auditors attempt to control the risk that material error in the financial statements will be undetected. Some auditors who have explicitly formulated a risk model use

\[ AR = IR \times CR \times AnR \times TD, \]

where \( AR \) = Audit risk; \( IR \) = Inherent risk — the risk that material error will occur in the absence of any internal accounting control; \( CR \) = Control risk — the risk that the system of internal accounting control will fail to prevent or detect material error should it occur; \( AnR \) = Analytical review risk; and \( TD \) = Test of details risk. The combination of the last two is referred to as detection risk — the risk that the auditor's procedures will fail to detect a material error.
that has not been prevented or detected by the system of control. As is obvious from these definitions, the model deals with materiality indirectly, but it is included in the assessment of risk.

The foregoing formulation (whether explicit or implicit) results in an appropriate assessment of audit risk only if each of the components is assessed reasonably well in relation to true financial statement materiality. Without making it explicit, the authors have assumed that this is the case.

It would seem to me that the product of IR and CR probably decreases as the number of locations increases. One million dollars of inventory all at one location, for example, can probably be misstated more easily than if the inventory is at 20 locations. Two issues are involved: (1) the number of locations where something would need to go wrong and (2) the relative significance of the inventory at each location.

Also, as the number of locations increases, I would expect that overall size tends to increase. Recent studies indicate that the materiality used to design audits increases at a decreasing rate. I believe that this decreasing rate is a function of client expectations and other factors rather than a "true" indication of materiality.

If this is the case, then today's auditors may have implicitly addressed the multilocation problem. Accordingly, caution should be exercised in the interpretation of the Aldersley-Lebl paper since the implication that today's auditors necessarily are underauditing in the multilocation situation does not follow.

In addition to the foregoing overall concern, the following specifics should be considered:

(1) The authors' concern about our literature being deficient in "determining what constitutes sufficient evidential matter ..." probably should not have been restricted to multilocation environments.

(2) The example in the third section probably should have specifically highlighted that one location had more inventory than the maximum level in the planning assumption. Also open to question is whether or not the additional observation should be added to the population, given that it could never have been selected in the sample.

(3) The approach to evaluation used in the third section is the same as the Stringer method as far as ranking and precision-adjusting the errors. However, it is not like the Stringer method in the treatment of understatements.

(4) The formula, in the fourth section, for identifying a fluctuation that warrants investigation is not a general formula. Rather it is a
specific formula for understatements. Also, the use of \( \gamma \) as a continuous function may result in answers that are more conservative than they need to be.

(5) Some thoughts on ways to improve the regression application:
   (a) A stratified example dealing with larger and smaller locations might help — seven of the nine base observations with observed values under 20,000 have negative residuals.
   (b) Observations 14 and 25 have large residuals — do we know something from last year’s audit?
   (c) With residuals in the audit period showing 2 to 1 negatives, has the model shifted since the base period?

(6) The point that errors found through the performance of analytical review procedures should be extrapolated to the population, based on the probability that the observation would not have been investigated, is very important. It would be obvious in a regression application that was designed to catch at least one if material error should exist. However, the logic applies equally to the judgmental analytical review involving, for example, a comparison of this year’s monthly or quarterly observations with last year’s amounts. There is more chance of investigating observations that are way out of line and less chance for those that match expectations. This difference in selection probabilities should be considered when the auditor relates the analytical review results to the population and the financial statements taken as a whole.

In conclusion, the Aldersley-Leslie paper certainly addresses an issue that needs further research attention so that we in the profession can better understand what we do and why we do it when audit effort is allocated in a multilocation environment.

NOTE

* I would like to thank the University of Illinois for sponsoring this audit symposium. It was nice to be back in Champaign-Urbana again. However, this time, I did have an advantage — I got to criticize the practitioner instead of being the practitioner that got criticized.
Estimation of Errors in Financial Statements

Detecting material errors in a client's financial statements is a major component of the external auditor's investigation leading to an opinion on financial statements. Our objectives herein are to (1) provide empirical evidence on errors discovered during year-end audits of manufacturing firms' accounts receivable and inventory and (2) determine the degree to which errors can be estimated given knowledge of general and account-specific company characteristics.

ERRORS IN FINANCIAL STATEMENTS AND AUDIT PROCEDURES

Audit Risk and Audit Planning

Audit risk (ARisk) is defined as the existence of an undetected material error in the financial statements. The audit risk model (see Statements on Auditing Standards (SAS) Nos. 39 and 47, AICPA 1984) provides a useful framework for discussion of error detection. Audit

Editors' Note: This paper was anonymously refereed prior to acceptance.
risk at the account level for a specific company can be thought of as a function of three "subrisks":

\[ ARisk = f(IR, IC, DR), \]

where \( IR \) is the probability that, given the nature of the company's operations, a material error will occur in account \( i \) (that is, the "inherent risk"); \( IC \) is the probability that the internal controls will not detect a material error entering, or being generated by, the client's accounting system (that is, the "control risk"); and \( DR \) is the probability that the selected audit procedures (analytical review procedures and tests of details) will not detect a material error, given that the error was not detected by the internal controls (that is, the "detection risk").

Although \( DR \) is conditional on \( IC, IR \) (the probability that a material error or irregularity exists in an account balance before consideration of the effectiveness of controls and the nature and intensity of the auditor's examination) is independent of \( IC \). The \( IR \) results from the direct and indirect effects of company characteristics that might lead to errors including the volatility of the firm's operations in a particular industry (product markets, and so on), the quality of management, characteristics of the company's information system and the entity's prevailing economic environment.\(^2\)

Estimating the probability of an undetected error for an account balance, given the \( IR \) and \( IC \) risks, is the initial step in planning an audit. The nature (compliance tests, analytical review procedures, and tests of details), extent (for example, the sample size), and timing (interim vs. year-end) of the audit procedures can be dependent on the perceived probability of a material error.

**RESEARCH ON FINANCIAL STATEMENT ERRORS**

Research on error magnitudes and probabilities may permit more refined and reliable statistical sampling procedures, given knowledge of the underlying error population distributions (Fesigne 1979; Johnson, Leitch, and Neter 1981). Viewing the emergence of errors as a stochastic process, information on this process could lead to more efficient data gathering by the auditor, and decision aids for sampling (see what follows). For example, given information on the probability of different error magnitudes at time \( t \), the auditor could concentrate on finding the largest number of errors, a range of error magnitudes and/or the very large errors.
Knowledge of error distribution properties also should be helpful in "fine tuning" analytical review procedures (Kinney 1978; 1979). Procedures such as trend comparisons and comparisons within a set of trial balance accounts at the same point in time, for example, implicitly include assumptions regarding the multivariate distribution of the account balances as well as the errors in those account balances. Finally, the design of audit decision aids, and the type of support for auditor judgment and decision making, should reflect knowledge of possible errors for individual accounts. For example, the present research describes error estimation models conditional on information about characteristics of a manufacturing company. A knowledge-based system could suggest a set of audit procedures based on the output of such estimation models and, thereby, enhance audit efficiency and perhaps, effectiveness.3

**Population Errors vs. Errors Detected by the Audit Process**

It is important to make a distinction between existing errors ("population errors") and errors actually detected by the (external) auditing examination ("detected errors"). The number and magnitude of detected errors will be a function of the number and magnitude of the population errors in the account, the effectiveness of the applicable internal controls in detecting and correcting errors, and the nature, extent, and timing of the portfolio of auditor's procedures. Or, equivalently, detected errors will be affected by inherent, control, and detection risks, respectively. The database used in this study consists of errors detected by the auditor during the audit examination, which include the joint effects of inherent, control, and detection risks.

If inherent risk is the same for a sample of audit clients, ineffective internal control is expected to provide for more frequent, and perhaps larger, errors in the accounting system. If an (accurate) assessment were to be made indicating that controls are relatively ineffective, the auditor should perform additional analytical review and tests of details, and consequently should detect a large number of errors (that is, since more errors were not corrected by the controls and more audit work was performed). Conversely, given reliance on very effective controls, the auditor should perform fewer tests and therefore, discover fewer errors (since the controls corrected more errors and less testing was performed by the external auditor).
METHODOLOGY

Overall Design

As discussed, a variety of company and situational factors could (directly and indirectly) determine the level of detected error in an account balance. Although one encounters informal suggestions in the academic literature concerning the variables that might be associated with errors, theoretical error-estimation models are not available.

Given the limited amount of theoretical and empirical work on such estimation models, we employed an exploratory approach herein. Specifically, judgments first were elicited from two experienced auditors concerning which variables, considered simultaneously, might be most associated with account misstatements. A list of possible variables was prepared, based upon these judgments (see Table 1). Subsequently, the variables were categorized based on their conceptual dimensions, and, based on correlations with detected errors, variables in each category were selected for further study (see Table 2).

Prior to being informed of the empirical results, each of two senior audit managers generated a list of the ten or fewer variables which they perceived to be most associated with inventory (INV) and accounts receivable (AR). They also were asked to explain why they selected the listed variables. The senior managers and one of the authors then met to discuss the reasoning behind the lists they developed and to decide upon a judgment-based model for empirical testing.

The following eleven variables constitute the initial judgment-based model (see Table 1). (The major categories for each variable correspond to the categories indicated in the Appendix.) (1) Category A. Quality of, and Situational Characteristics for, Accounting Personnel (ATPRESS, AKNOW); (2) Category B. Nature and Quality of Information Systems (CINFOSYS); (3) Category C. General Controls (GENCONT and *INCONT); (4) Category E. Degree of Emphasis by Management on Accounting Earnings (EARNPROE, EARNCOMP); (5) Several account-specific, accounting complexity indications (*JUDGE, *COMP, and *MATER); and (6) Error information from the previous year (AR_{t-1} or INV_{t-1}).

The magnitude of any detected error in the previous year (AR_{t-1}, INV_{t-1}) was included in the model based upon prior research (Kinney 1979). The model also included a measure of internal control effectiveness (INCONT) in addition to the general controls measure (GENCONT).
Table 1. Judgment-based Estimation Model

The following general and account-specific (indicated with a "**") variables were selected from the variables in the Appendix.

A. Quality of, and Situational Characteristics for, Accounting Personnel
1. ATPRESS — Indicate the degree of time pressure on the entity’s accounting personnel (i.e., Controller and staff)
2. AKNOW — Rate the level of knowledge of the entity’s accounting personnel (i.e., Controller and staff) in terms of their awareness and understanding of accounting principles and practices and how to apply them. Consider college degrees held, training courses attended, and your observations of the personnel.

B. Nature and Quality of Information Systems
3. CINFOSYS — To what degree did the entity’s financial information system(s) change in the last fiscal year

C. General Controls
4. GENCONT — Rate the overall level of general controls, including the potential for management override. Consider factors such as organizational structure, documentation policies, existence of budgets and comparison of budgets to actual results, and existence of an internal audit department.
5. *INCONT — Rate the effectiveness of the internal controls affecting the account.

D. Degree of Emphasis on Accounting Earnings
6. EARNPROE — Rate (Executive) management’s emphasis on meeting earnings projections.
7. EARNCOMP — Evaluate the entity’s most recent net earnings relative to the entity’s net earnings in the prior period.

Account-specific, accounting complexity indications
8. *JUDGE — Rate the degree to which judgment (including estimates) was required in arriving at the entries to the account.
9. *COMP — Rate the degree of complexity underlying the entries made to the account. For example, summarizing cash receipts usually is not complex, whereas calculating the liability and expense related to federal income taxes is often highly complex.
10. *MATER — Rate the materiality of the account balance.

Previous error information
11. Error information from the previous year (AR_{t-1}, INV_{t-1}).

After testing the judgment model, we considered additional variables (see the Appendix for the list and Table 2 for the data relationships). Several models were investigated. The final models were selected based on their explanatory power and understandability.
Table 2. Pearson Correlations for Errors and the Variables

<table>
<thead>
<tr>
<th>Panel A — Inventory (INV.) Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>CENTOPER  OUTSIDER  OUTSTURN  ALLEGRE  ALLEGSE  MCOMPETE</td>
</tr>
<tr>
<td>-.0784 .0330 .0935 -.3941 -.4810 .2518</td>
</tr>
<tr>
<td>( .79) ( .79) ( .61) ( .78) ( .67) ( .79)</td>
</tr>
<tr>
<td>P = .246 P = .386 P = .227 P = .001 P = .001 P = .013</td>
</tr>
<tr>
<td>MCOMPETM AGGRISKE AGGRISKM EXPERE EXPERM EARNPROE</td>
</tr>
<tr>
<td>.2663 .2948 .2873 .2006 .2092 .0724</td>
</tr>
<tr>
<td>( .77) ( .79) ( .74) ( .79) ( .77) ( .79)</td>
</tr>
<tr>
<td>P = .010 P = .004 P = .007 P = .038 P = .054 P = .263</td>
</tr>
<tr>
<td>EARNPROM COMPENE COMPENM MANTURNO AIINTEG AKNOW</td>
</tr>
<tr>
<td>.0639 .1534 .0661 .2661 .3792 .1781</td>
</tr>
<tr>
<td>( .76) ( .78) ( .75) ( .79) ( .79) ( .79)</td>
</tr>
<tr>
<td>P = .292 P = .090 P = .285 P = .009 P = .001 P = .058</td>
</tr>
<tr>
<td>AXPERM ACONSC ATPRESS ATTURN CINFOSYS AUTOMAT</td>
</tr>
<tr>
<td>.1628 .4741 .0326 .1898 .1219 .0405</td>
</tr>
<tr>
<td>( .79) ( .79) ( .79) ( .79) ( .79) ( .79)</td>
</tr>
<tr>
<td>P = .076 P = .001 P = .388 P = .047 P = .142 P = .361</td>
</tr>
<tr>
<td>GENCONT MLLETTER MLRESPON LIQUID LIQUIDC COMPLAON</td>
</tr>
<tr>
<td>.0745 .3825 .1722 .1755 .1381 .2405</td>
</tr>
<tr>
<td>( .79) ( .79) ( .70) ( .79) ( .79) ( .79)</td>
</tr>
<tr>
<td>P = .257 P = .001 P = .077 P = .061 P = .112 P = .040</td>
</tr>
<tr>
<td>RETURNA RESOPER RELEARN BUSOUT EARNCOMP TECHPRO</td>
</tr>
<tr>
<td>.1405 -.0942 .1347 .1490 .0660 .0733</td>
</tr>
<tr>
<td>( .79) ( .78) ( .72) ( .78) ( .79) ( .79)</td>
</tr>
<tr>
<td>P = .108 P = .206 P = .130 P = .096 P = .282 P = .260</td>
</tr>
<tr>
<td>TECHPSER CONTPROI INDCOMP REGCOMP INDCRHG IndoK</td>
</tr>
<tr>
<td>.1122 .1416 .1139 -.1723 .2062 .0894</td>
</tr>
<tr>
<td>( .78) ( .78) ( .79) ( .79) ( .77) ( .75)</td>
</tr>
<tr>
<td>P = .164 P = .108 P = .159 P = .064 P = .040 P = .223</td>
</tr>
<tr>
<td>INTRATE INFLAT EXCHGR POLENE INJUDG INCOMP</td>
</tr>
<tr>
<td>-.0048 .0959 -.0070 .1433 .0964 .1584</td>
</tr>
<tr>
<td>( .78) ( .79) ( .77) ( .78) ( .75) ( .76)</td>
</tr>
<tr>
<td>P = .483 P = .200 P = .476 P = .105 P = .204 P = .089</td>
</tr>
<tr>
<td>ININCONT INFRAUD INMATER INMUT INTRANS INRELAT</td>
</tr>
<tr>
<td>.1918 .0952 .0870 .0583 .3600 .1410</td>
</tr>
<tr>
<td>( .75) ( .78) ( .76) ( .76) ( .76) ( .76)</td>
</tr>
<tr>
<td>P = .050 P = .207 P = .227 P = .308 P = .001 P = .112</td>
</tr>
<tr>
<td>INSEAS -.0954</td>
</tr>
<tr>
<td>( .76)</td>
</tr>
<tr>
<td>P = .206</td>
</tr>
</tbody>
</table>
### Table 2. Continued

**Panel B — Accounts Receivable (AR) Correlations**

<table>
<thead>
<tr>
<th>CENTOPER</th>
<th>OUTSIDER</th>
<th>OUTSTURN</th>
<th>ALLEGFRE</th>
<th>ALLEGSER</th>
<th>MCOMPETM</th>
<th>AGGRISKF</th>
<th>AGGRISKM</th>
<th>EXPERE</th>
<th>EXPERM</th>
<th>EARNPROE</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.0435</td>
<td>0.0135</td>
<td>0.1047</td>
<td>-0.357</td>
<td>-0.0226</td>
<td>0.0940</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(79)</td>
<td>(79)</td>
<td>(61)</td>
<td>(78)</td>
<td>(67)</td>
<td>(79)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P = .352</td>
<td>P = .453</td>
<td>P = .211</td>
<td>P = .378</td>
<td>P = .428</td>
<td>P = .205</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MCOMPETM</th>
<th>AGGRISKF</th>
<th>AGGRISKM</th>
<th>EXPERE</th>
<th>EXPERM</th>
<th>EARNPROE</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1200</td>
<td>0.0644</td>
<td>0.0077</td>
<td>0.1255</td>
<td>0.1002</td>
<td>0.0180</td>
</tr>
<tr>
<td>(77)</td>
<td>(79)</td>
<td>(74)</td>
<td>(79)</td>
<td>(77)</td>
<td>(79)</td>
</tr>
<tr>
<td>P = .149</td>
<td>P = .286</td>
<td>P = .474</td>
<td>P = .141</td>
<td>P = .193</td>
<td>P = .437</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FARNPROE</th>
<th>COMPENE</th>
<th>COMPENM</th>
<th>MANTURNO</th>
<th>AINTEG</th>
<th>AKNOW</th>
</tr>
</thead>
<tbody>
<tr>
<td>.0318</td>
<td>0.0884</td>
<td>0.1284</td>
<td>.2530</td>
<td>.2028</td>
<td>.0192</td>
</tr>
<tr>
<td>(76)</td>
<td>(78)</td>
<td>(76)</td>
<td>(79)</td>
<td>(79)</td>
<td>(79)</td>
</tr>
<tr>
<td>P = .393</td>
<td>P = .221</td>
<td>P = .134</td>
<td>P = .012</td>
<td>P = .037</td>
<td>P = .433</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AEXPER</th>
<th>ACONSC</th>
<th>ATPRESS</th>
<th>ATURN</th>
<th>CINFOSYS</th>
<th>AUTOMAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>.0846</td>
<td>0.2946</td>
<td>0.0476</td>
<td>0.2726</td>
<td>0.0682</td>
<td>0.2241</td>
</tr>
<tr>
<td>P = .229</td>
<td>P = .004</td>
<td>P = .339</td>
<td>P = .008</td>
<td>P = .275</td>
<td>P = .024</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>GENCONT</th>
<th>MLLETTER</th>
<th>MLRESPON</th>
<th>LIQUID</th>
<th>LIQUIDC</th>
<th>COMPOAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.0952</td>
<td>0.1444</td>
<td>-0.1235</td>
<td>0.1475</td>
<td>0.1652</td>
<td>0.1276</td>
</tr>
<tr>
<td>(79)</td>
<td>(79)</td>
<td>(70)</td>
<td>(79)</td>
<td>(79)</td>
<td>(70)</td>
</tr>
<tr>
<td>P = .292</td>
<td>P = .102</td>
<td>P = .154</td>
<td>P = .097</td>
<td>P = .073</td>
<td>P = .179</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RETURNA</th>
<th>RESOPER</th>
<th>RELEARN</th>
<th>BUSOUT</th>
<th>EARNCOMP</th>
<th>TECHPRO</th>
</tr>
</thead>
<tbody>
<tr>
<td>.0331</td>
<td>-0.0374</td>
<td>0.0914</td>
<td>-0.0860</td>
<td>0.1120</td>
<td>0.0630</td>
</tr>
<tr>
<td>(79)</td>
<td>(78)</td>
<td>(72)</td>
<td>(78)</td>
<td>(79)</td>
<td>(79)</td>
</tr>
<tr>
<td>P = .386</td>
<td>P = .375</td>
<td>P = .222</td>
<td>P = .227</td>
<td>P = .163</td>
<td>P = .291</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TECHPSER</th>
<th>CONTROI</th>
<th>INDCOMP</th>
<th>REGCOMP</th>
<th>INDCGH</th>
<th>INDPPOF</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1893</td>
<td>0.1126</td>
<td>0.0840</td>
<td>-0.1438</td>
<td>0.0808</td>
<td>0.0662</td>
</tr>
<tr>
<td>(78)</td>
<td>(78)</td>
<td>(79)</td>
<td>(79)</td>
<td>(77)</td>
<td>(78)</td>
</tr>
<tr>
<td>P = .048</td>
<td>P = .163</td>
<td>P = .231</td>
<td>P = .103</td>
<td>P = .243</td>
<td>P = .286</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>INTRATE</th>
<th>INFLAT</th>
<th>EXCHGR</th>
<th>POLENVE</th>
<th>INJUDG</th>
<th>INCOMP</th>
</tr>
</thead>
<tbody>
<tr>
<td>.0762</td>
<td>-0.1245</td>
<td>0.1725</td>
<td>0.1179</td>
<td>-0.0552</td>
<td>-.0035</td>
</tr>
<tr>
<td>(78)</td>
<td>(79)</td>
<td>(77)</td>
<td>(78)</td>
<td>(79)</td>
<td>(76)</td>
</tr>
<tr>
<td>P = .254</td>
<td>P = .137</td>
<td>P = .067</td>
<td>P = .152</td>
<td>P = .315</td>
<td>P = .488</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ARINGSA</th>
<th>ARRAUFD</th>
<th>ARMATEK</th>
<th>ARMUTF</th>
<th>ARTRANS</th>
<th>ARRELAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>.1548</td>
<td>0.0062</td>
<td>-0.1057</td>
<td>0.0506</td>
<td>0.1246</td>
<td>0.0436</td>
</tr>
<tr>
<td>(75)</td>
<td>(74)</td>
<td>(74)</td>
<td>(74)</td>
<td>(74)</td>
<td>(74)</td>
</tr>
<tr>
<td>P = .128</td>
<td>P = .479</td>
<td>P = .183</td>
<td>P = .307</td>
<td>P = .145</td>
<td>P = .356</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ARSEAS</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>.1126</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(74)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P = .170</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Data

We utilized a database (PMM Risk Database) developed at Peat, Marwick, Mitchell & Co. in conducting the study. This database consists of information provided by audit managers and in-charge seniors which was taken from their year-end audit work papers. It contains detailed information on actual audit clients including general and specific company characteristics, information for two consecutive years on individual financial statement line items and errors (that is, booked and waived audit differences — differences between client and auditor indications of account balances). The data were obtained in late 1982 for 82 manufacturing firms having the following financial characteristics (in terms of fiscal 1981 data):

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sales</td>
<td>$83,247,500</td>
<td>$19,694,000</td>
</tr>
<tr>
<td>Earnings before taxes</td>
<td>3,583,100</td>
<td>772,200</td>
</tr>
<tr>
<td>Net earnings</td>
<td>2,346,300</td>
<td>431,700</td>
</tr>
<tr>
<td>Total assets</td>
<td>58,129,300</td>
<td>13,317,000</td>
</tr>
</tbody>
</table>

In constructing the database, the error magnitude was computed as the absolute value of the net dollar error for trade AR and INV including errors in the associated allowance accounts. Additionally, some errors were transaction-processing-oriented (for example, computational mistakes, omissions, or classification errors) while others were more subjective (for example, incorrect asset valuations such as an inappropriate net realizable value for inventory items or an inappropriate allowance for doubtful accounts). Incorrect allocations of cost variances also were included as errors. Therefore, the net per account error may include transaction processing, valuation, and cost allocation errors. These errors were detected as a result of analytical review procedures and other substantive audit procedures such as analyses of the quantities and prices of purchases, confirmation results, and cut-off procedures. Since, however, the PMM Risk Database consists only of detected errors, the effects of inherent, control, and detection risks are jointly impounded.

RESULTS

Evidence on Distribution Characteristics of Errors

Table 3 provides a number of statistics on the distributions of errors across the 82 companies, including relative likelihood and magnitude information; Table 4 provides similar data on the variables
Table 3. Characteristics of the Error Distributions

Panel A — Signed values of the (net) errors divided by audit gauge (x 100.0)

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>Mode</th>
<th>Kurtosis</th>
<th>Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR_i</td>
<td>-15.772</td>
<td>-3.000</td>
<td>0</td>
<td>11.287</td>
<td>0.942</td>
</tr>
<tr>
<td>Std dev</td>
<td>49.229</td>
<td>11.287</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>418.000</td>
<td>-173.000</td>
<td>Maximum</td>
<td>245.000</td>
<td></td>
</tr>
<tr>
<td>Valid cases</td>
<td>79</td>
<td>Missing cases</td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>Mode</th>
<th>Kurtosis</th>
<th>Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR_{i-1}</td>
<td>10.620</td>
<td>0.083</td>
<td>0</td>
<td>39.383</td>
<td>5.671</td>
</tr>
<tr>
<td>Std dev</td>
<td>58.300</td>
<td>12.934</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>412.000</td>
<td>-92.000</td>
<td>Maximum</td>
<td>440.000</td>
<td></td>
</tr>
<tr>
<td>Valid cases</td>
<td>79</td>
<td>Missing cases</td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>Mode</th>
<th>Kurtosis</th>
<th>Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>INV_i</td>
<td>-1.595</td>
<td>2.35</td>
<td>0</td>
<td>-1.498</td>
<td>-1.892</td>
</tr>
<tr>
<td>Std dev</td>
<td>54.838</td>
<td>12.934</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>508.000</td>
<td>-295.000</td>
<td>Maximum</td>
<td>213.000</td>
<td></td>
</tr>
<tr>
<td>Valid cases</td>
<td>79</td>
<td>Missing cases</td>
<td>4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Panel B — Absolute values of the (net) errors divided by audit gauge (x 100.0)

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>Mode</th>
<th>Kurtosis</th>
<th>Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR_i</td>
<td>26.304</td>
<td>7.250</td>
<td>0</td>
<td>8.539</td>
<td>2.705</td>
</tr>
<tr>
<td>Std dev</td>
<td>44.437</td>
<td>8.539</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>245.000</td>
<td>0</td>
<td>Maximum</td>
<td>245.000</td>
<td></td>
</tr>
<tr>
<td>Valid cases</td>
<td>79</td>
<td>Missing cases</td>
<td>3</td>
<td>n of zeros = 30</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>Mode</th>
<th>Kurtosis</th>
<th>Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR_{i-1}</td>
<td>17.557</td>
<td>0.231</td>
<td>0</td>
<td>41.488</td>
<td>5.992</td>
</tr>
<tr>
<td>Std dev</td>
<td>56.577</td>
<td>41.488</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>440.000</td>
<td>0</td>
<td>Maximum</td>
<td>440.000</td>
<td></td>
</tr>
<tr>
<td>Valid cases</td>
<td>79</td>
<td>Missing cases</td>
<td>3</td>
<td>n of zeros = 54</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>Mode</th>
<th>Kurtosis</th>
<th>Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>INV_i</td>
<td>27.873</td>
<td>10.333</td>
<td>0</td>
<td>15.390</td>
<td>3.543</td>
</tr>
<tr>
<td>Std dev</td>
<td>47.147</td>
<td>15.390</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>295.000</td>
<td>0</td>
<td>Maximum</td>
<td>295.000</td>
<td></td>
</tr>
<tr>
<td>Valid cases</td>
<td>79</td>
<td>Missing cases</td>
<td>3</td>
<td>n of zeros = 17</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>Mode</th>
<th>Kurtosis</th>
<th>Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>INV_{i-1}</td>
<td>32.372</td>
<td>5.00</td>
<td>0</td>
<td>4.838</td>
<td>2.307</td>
</tr>
<tr>
<td>Std dev</td>
<td>57.740</td>
<td>4.838</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>254.000</td>
<td>0</td>
<td>Maximum</td>
<td>254.000</td>
<td></td>
</tr>
<tr>
<td>Valid cases</td>
<td>78</td>
<td>Missing cases</td>
<td>4</td>
<td>n of zeros = 39</td>
<td></td>
</tr>
</tbody>
</table>

* The errors were divided by a measure of client size ("audit gauge") to be comparable in relative terms across firms (see note 6).
<table>
<thead>
<tr>
<th>Variable</th>
<th>No. of Obs.</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CENTOPER</td>
<td>81</td>
<td>2.2963</td>
<td>1.7135</td>
</tr>
<tr>
<td>OUTSIDER</td>
<td>81</td>
<td>5.4938</td>
<td>1.7473</td>
</tr>
<tr>
<td>OUTSTURN</td>
<td>61</td>
<td>1.5738</td>
<td>0.6444</td>
</tr>
<tr>
<td>ALLEGFRE</td>
<td>80</td>
<td>6.4000</td>
<td>1.1649</td>
</tr>
<tr>
<td>ALLEGSER</td>
<td>68</td>
<td>6.5000</td>
<td>1.1395</td>
</tr>
<tr>
<td>MCOMPETE</td>
<td>81</td>
<td>2.8395</td>
<td>1.4094</td>
</tr>
<tr>
<td>MCOMPETM</td>
<td>79</td>
<td>3.4684</td>
<td>1.2283</td>
</tr>
<tr>
<td>AGGRISEK</td>
<td>81</td>
<td>3.5309</td>
<td>1.4325</td>
</tr>
<tr>
<td>AGGRISKM</td>
<td>76</td>
<td>3.0526</td>
<td>1.2847</td>
</tr>
<tr>
<td>EXPER</td>
<td>81</td>
<td>2.1481</td>
<td>0.9501</td>
</tr>
<tr>
<td>EXPERM</td>
<td>79</td>
<td>2.9114</td>
<td>1.1000</td>
</tr>
<tr>
<td>EARNPROE</td>
<td>81</td>
<td>4.6790</td>
<td>1.4733</td>
</tr>
<tr>
<td>EARNPRM</td>
<td>78</td>
<td>4.3513</td>
<td>1.3353</td>
</tr>
<tr>
<td>COMPENS1</td>
<td>80</td>
<td>3.8125</td>
<td>1.6311</td>
</tr>
<tr>
<td>COMPENS2</td>
<td>78</td>
<td>3.5385</td>
<td>1.3835</td>
</tr>
<tr>
<td>MANTURNO</td>
<td>81</td>
<td>2.1358</td>
<td>1.2821</td>
</tr>
<tr>
<td>AINTEG</td>
<td>81</td>
<td>2.1728</td>
<td>1.1268</td>
</tr>
<tr>
<td>AKNOW</td>
<td>81</td>
<td>3.6296</td>
<td>1.5914</td>
</tr>
<tr>
<td>AEXPER</td>
<td>81</td>
<td>3.0270</td>
<td>1.1118</td>
</tr>
<tr>
<td>ACONSC</td>
<td>81</td>
<td>2.2840</td>
<td>1.0155</td>
</tr>
<tr>
<td>ATPRESS</td>
<td>81</td>
<td>3.9136</td>
<td>1.3982</td>
</tr>
<tr>
<td>ATURN</td>
<td>81</td>
<td>2.6790</td>
<td>1.3493</td>
</tr>
<tr>
<td>CINFOSYS</td>
<td>81</td>
<td>2.5679</td>
<td>1.4827</td>
</tr>
<tr>
<td>AUTOMAT</td>
<td>81</td>
<td>4.3827</td>
<td>1.7928</td>
</tr>
<tr>
<td>GENCONT</td>
<td>81</td>
<td>3.7407</td>
<td>1.2921</td>
</tr>
<tr>
<td>MLETTRE</td>
<td>81</td>
<td>3.0570</td>
<td>1.5447</td>
</tr>
<tr>
<td>MLRESPON</td>
<td>72</td>
<td>3.2639</td>
<td>1.4339</td>
</tr>
<tr>
<td>LIQUID</td>
<td>81</td>
<td>3.4938</td>
<td>1.7967</td>
</tr>
<tr>
<td>LIQUIDC</td>
<td>81</td>
<td>3.9883</td>
<td>1.2683</td>
</tr>
<tr>
<td>COMPLANO</td>
<td>56</td>
<td>2.9286</td>
<td>1.8960</td>
</tr>
<tr>
<td>RETURN</td>
<td>81</td>
<td>4.0864</td>
<td>1.9950</td>
</tr>
<tr>
<td>RESOPER</td>
<td>80</td>
<td>5.6750</td>
<td>1.3099</td>
</tr>
<tr>
<td>RELEARN</td>
<td>73</td>
<td>3.9863</td>
<td>1.5229</td>
</tr>
<tr>
<td>BUSOUT</td>
<td>80</td>
<td>4.0000</td>
<td>1.3962</td>
</tr>
<tr>
<td>EARNCOMP</td>
<td>81</td>
<td>4.1235</td>
<td>1.7775</td>
</tr>
<tr>
<td>TECHPRO</td>
<td>81</td>
<td>4.4815</td>
<td>1.3974</td>
</tr>
<tr>
<td>TECHPSER</td>
<td>80</td>
<td>4.1500</td>
<td>1.6512</td>
</tr>
<tr>
<td>CONTPROI</td>
<td>80</td>
<td>3.2125</td>
<td>1.7552</td>
</tr>
<tr>
<td>INDCOMP</td>
<td>81</td>
<td>4.9259</td>
<td>1.3208</td>
</tr>
<tr>
<td>REGCOMP</td>
<td>81</td>
<td>4.5309</td>
<td>1.5090</td>
</tr>
<tr>
<td>INDCIH</td>
<td>79</td>
<td>3.6329</td>
<td>1.3222</td>
</tr>
<tr>
<td>INDPERS</td>
<td>77</td>
<td>4.0909</td>
<td>1.3687</td>
</tr>
<tr>
<td>INTRATE</td>
<td>80</td>
<td>4.2125</td>
<td>1.8050</td>
</tr>
<tr>
<td>INFLAT</td>
<td>81</td>
<td>4.4815</td>
<td>1.3239</td>
</tr>
<tr>
<td>EXCHGR</td>
<td>79</td>
<td>2.4430</td>
<td>1.6152</td>
</tr>
<tr>
<td>POLLENVE</td>
<td>80</td>
<td>2.9875</td>
<td>1.5791</td>
</tr>
</tbody>
</table>
Table 4 (continued)

<table>
<thead>
<tr>
<th>Predictor</th>
<th>No. of Obs.</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>INJUDG</td>
<td>78</td>
<td>3.8846</td>
<td>1.4681</td>
</tr>
<tr>
<td>INCOMP</td>
<td>78</td>
<td>4.3205</td>
<td>1.5583</td>
</tr>
<tr>
<td>ININCONT</td>
<td>77</td>
<td>3.6623</td>
<td>1.3535</td>
</tr>
<tr>
<td>INFRAUD</td>
<td>78</td>
<td>3.4231</td>
<td>1.2329</td>
</tr>
<tr>
<td>INMATER</td>
<td>78</td>
<td>5.6026</td>
<td>1.6852</td>
</tr>
<tr>
<td>INMUT</td>
<td>78</td>
<td>3.1538</td>
<td>1.7138</td>
</tr>
<tr>
<td>INTRANS</td>
<td>78</td>
<td>2.4359</td>
<td>1.3919</td>
</tr>
<tr>
<td>INRELS</td>
<td>78</td>
<td>2.3128</td>
<td>1.2766</td>
</tr>
<tr>
<td>INSEAS</td>
<td>78</td>
<td>3.2436</td>
<td>1.8032</td>
</tr>
<tr>
<td>ARJUDG</td>
<td>81</td>
<td>2.7654</td>
<td>1.2676</td>
</tr>
<tr>
<td>ARCOMP</td>
<td>81</td>
<td>2.2840</td>
<td>1.0276</td>
</tr>
<tr>
<td>ARINCONT</td>
<td>75</td>
<td>3.3067</td>
<td>1.4702</td>
</tr>
<tr>
<td>ARFRAUD</td>
<td>76</td>
<td>2.5263</td>
<td>1.2380</td>
</tr>
<tr>
<td>ARMATER</td>
<td>76</td>
<td>5.5395</td>
<td>1.5701</td>
</tr>
<tr>
<td>ARMUT</td>
<td>76</td>
<td>3.8642</td>
<td>1.7904</td>
</tr>
<tr>
<td>ARTRANS</td>
<td>76</td>
<td>2.0526</td>
<td>1.0314</td>
</tr>
<tr>
<td>ARRELAT</td>
<td>76</td>
<td>2.2105</td>
<td>1.2787</td>
</tr>
<tr>
<td>ARSEAS</td>
<td>76</td>
<td>3.1974</td>
<td>1.7283</td>
</tr>
</tbody>
</table>

described in the Appendix. The net dollar error for INV and AR was divided by a measure of the size of the company ("audit gauge") to adjust for size differences across firms, and multiplied by 100 to be in percentages of audit gauge.5

Panel A of Table 3 describes the signed errors while Panel B indicates the absolute values of the errors. The absolute values of the errors as a percentage of gauge were used to develop the estimation models. This approach assumes a symmetric loss function (that is, overstatement and understatement of the account balance at a point in time are equally important).6

Error distributions are characterized by several "zero" observations (no net detected error in the account at year-end) and a few very large errors (see Table 3, Panel B and Charts 1 and 2). The distributions are unimodal, with substantial positive skewness, and the null hypothesis of normality is rejected for both AR, and INV, (also AR_{t-1} and INV_{t-1}) with z-statistics from Kolmogorov-Smirnov tests of normality of at least 2.4 (p<.001).5

Concerning model development, relative to a symmetric (for example, normal) distribution, two major problems exist in the data: (1) relative to the mean, the distributions are essentially truncated by the no-error, "zero" observations and (2) the few large errors
Pearson correlation = .40456 (p=.00014).
A plotted "9" indicates 9 or more observations — 76 observations plotted.
Chart 2. Scatterplot of AR Errors Current Year vs. AR Errors Preceding Year

Pearson correlation = .42547 (p=.00006).
A plotted "9" indicates 9 or more observations — 77 observations plotted.
result in substantial positive skewness. Transformation of INV, and INV,-1 (and both AR, and AR,-1) was attempted, therefore, to make the distributions more consistent with the normality (or at least symmetric, unimodal) assumption.

Although logarithmic transformation diminished the effects of both of these departures from normality, normal probability plots and z-statistics from Kolmogorov-Smirnov tests continued to indicate non-normal error distributions. Consequently, the significance level results reported herein must be viewed as approximations.

Explaining Inventory Error

Correlations between the variables in the judgment model and the INV errors are reported in Table 5, Panel A. Such correlations generally are low except for INV,-1 (that is, the correlation between errors detected this year (INV,) and errors detected last year (INV,-1) is 0.417 (p<.001)). Based on their correlations with INV,, four other variables (CINFOSYS, AKNOW, ININCONT, and INCOMP) were identified as potentially meaningful contributors to the model (see Table 5, Panel A).

Regression results for the judgment model are reported in Panel B of Table 5. Notice that when an F-statistic of 1.80 (approximate p=.15) is required for inclusion in the stepwise regression, only INV,-1 is significant. Additionally, when information on errors in the previous period is available, it overwhelms the other variables in terms of fit. The model is significant (F=14.75, p<.001), with an adjusted $R^2$ of 0.1629, although considerable variance in INV remains unexplained. ININCONT would have been the next variable to enter the model with an F-statistic of 1.67 (p=.20) and the adjusted $R^2$ would have increased from 0.1629 to 0.1702.

Next, the error/variable correlations of the other variables were considered as a means of improving the model. Relatively large correlations are reported for variables in four conceptual categories (see Table 2, Panel A and the Appendix): (1) Category A, the quality of, and situational characteristics for, accounting personnel (AINTEG, AKNOW, AEXPER, ACONSC, AND ATURN); (2) Category B, the nature and quality of company’s information systems (MLETTER); (3) Category D, the competence, aggressiveness and experience of executive, senior and middle management (MCOMPETE, MPCOMPETM, AGGRISKE, AGGRISKM, EXPERE, EXPERM AND MANTURNO); and (4) the number of unusual inventory transactions (INTRANS).

Starting with the model described in Table 5, Panel B, the variables
Table 5. Models for Inventory Errors

Panel A — Correlation Matrix for Judgment Model

<table>
<thead>
<tr>
<th></th>
<th>INV_{t-1}</th>
<th>EARNPROE</th>
<th>EARNCOMP</th>
<th>CINFOSYS</th>
<th>ATPRESS</th>
<th>AKNOW</th>
<th>GENCONT</th>
<th>ININCONT</th>
<th>INJUDG</th>
<th>INCOMP</th>
<th>INMATER</th>
<th>GENCONT</th>
<th>ININCONT</th>
<th>INJUDG</th>
<th>INCOMP</th>
<th>INMATER</th>
</tr>
</thead>
<tbody>
<tr>
<td>INV_{t-1}</td>
<td>.41722</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EARNPROE</td>
<td></td>
<td>-.14374</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EARNCOMP</td>
<td></td>
<td></td>
<td>.12672</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CINFOSYS</td>
<td></td>
<td></td>
<td></td>
<td>.15655</td>
<td>-.02456</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATPRESS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.21791</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AKNOW</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.33035</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GENCONT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.11728</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ININCONT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.32358</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INJUDG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.6582</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INCOMP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.17838</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INMATER</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.30802</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INV_{t-1}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EARNPROE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EARNCOMP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CINFOSYS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATPRESS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AKNOW</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GENCONT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ININCONT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INJUDG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INCOMP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INMATER</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Panel B — Test of Judgment Model

| Mean INV_{t} | INV_{t} | 29.37500 | ANOVA
|-------------|--------|---------|--------
| Std. Dev. (INV_{t}) | 48.86672 | DF | Sum Squares | Mean Sq | F |
| Multiple R | .4172 | Regression | 1 | 29512.744 | 29512.744 | 14.753 |
| R Square | .1741 | Residual | 70 | 140032.131 | 2000.459 | SIG .000 |
| Std Dev | 44.7265 | | | | |
| Adj R Square | .1623 | Variable | B | S.E. B | F | Sig. | Beta |
| INV_{t-1} | .350 | INV_{t-1} | .091 | 14.753 | .000 | .41722 |
| Constant | 17.756 | Constant | 6.077 | 8.536 | .005 | |
### Table 5. Continued

**Panel C — Revised INV Error Model**

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>S.E. B</th>
<th>F</th>
<th>Sig.</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>INV</td>
<td>.227</td>
<td>.083</td>
<td>7.404</td>
<td>.008</td>
<td>.27096</td>
</tr>
<tr>
<td>ACONSC</td>
<td>17.386</td>
<td>4.874</td>
<td>12.725</td>
<td>.001</td>
<td>.36199</td>
</tr>
<tr>
<td>INTRANS</td>
<td>9.572</td>
<td>3.293</td>
<td>8.451</td>
<td>.005</td>
<td>.27848</td>
</tr>
<tr>
<td>Constant</td>
<td>-40.421</td>
<td>12.552</td>
<td>10.371</td>
<td>.002</td>
<td></td>
</tr>
</tbody>
</table>

**ANOVA**

<table>
<thead>
<tr>
<th>DF</th>
<th>Sum Squares</th>
<th>Mean Sq.</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>68406.281</td>
<td>22802.094</td>
<td>15.331</td>
</tr>
<tr>
<td>68</td>
<td>101138.594</td>
<td>1487.332</td>
<td></td>
</tr>
</tbody>
</table>

**Panel D — Final INV Error Model**

<table>
<thead>
<tr>
<th>Variable</th>
<th>B</th>
<th>S.E. B</th>
<th>F</th>
<th>Sig.</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>INV</td>
<td>.244</td>
<td>.083</td>
<td>8.441</td>
<td>.005</td>
<td>.29724</td>
</tr>
<tr>
<td>INTRANS</td>
<td>7.553</td>
<td>3.426</td>
<td>4.859</td>
<td>.091</td>
<td>.22002</td>
</tr>
<tr>
<td>ACONSC</td>
<td>11.152</td>
<td>5.119</td>
<td>4.746</td>
<td>.033</td>
<td>.23499</td>
</tr>
<tr>
<td>MLETTER</td>
<td>5.764</td>
<td>3.022</td>
<td>3.047</td>
<td>.086</td>
<td>.18175</td>
</tr>
<tr>
<td>AGGRISKE</td>
<td>3.738</td>
<td>3.361</td>
<td>2.915</td>
<td>.093</td>
<td>.17108</td>
</tr>
<tr>
<td>Constant</td>
<td>-60.412</td>
<td>15.257</td>
<td>15.679</td>
<td>.000</td>
<td></td>
</tr>
</tbody>
</table>

**ANOVA**

<table>
<thead>
<tr>
<th>DF</th>
<th>Sum Squares</th>
<th>Mean Sq.</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>73668.363</td>
<td>14733.727</td>
<td>10.105</td>
</tr>
<tr>
<td>65</td>
<td>94778.660</td>
<td>1458.133</td>
<td></td>
</tr>
</tbody>
</table>
dealing with the quality of the accounting staff and INTRANS were considered as possible significant explanatory variables in a stepwise regression. The results are reported in Panel C.

The revised model indicated in Table 5, Panel C reveals the considerable explanatory power of (1) the quality of the accounting staff dimension via the ACONSC variable and (2) the number of unusual inventory transactions (INTRANS). Both variables are statistically significant in explaining INV, ($p<.01$). Notice that the revised model explains considerably more of the variance of the dependent variable INV, with the adjusted $R^2$ increasing from 0.1623 to 0.3772. The three variables in the model are each statistically significant and, with an $F$-statistic of 1.80 required to enter the model, none of the other variables were entered. One possible reason for the other variables not adding significantly to the model is the high intercorrelation among the variables; for example, the correlation between ACONSC and AINTEG (both of which are highly correlated with INV) is 0.681.

In addition to the quality of the entity's accounting staff, the other major categories of variables with relatively high pairwise INV, correlations are the quality of management variables (MCOMPETE, MCOMPETM, AGGRISKE, AGGRISKM, EXPERE, EXPERM, AND MANTURNO) and the quality of the entity's information and control systems (MLETTER). Thus, the variables in the previous model (Panel C), the management variables, and MLETTER were included as input into a stepwise regression and the resulting model is indicated in Table 5, Panel D.

Additional improvement of the model is indicated by the increased adjusted $R^2$ (from 0.3772 to 0.3941). Further, the model is statistically significant ($F=10.105$, $p<.001$) and each of the five explanatory variables is significant at the $p=.10$ level (however the foregoing comments on the lack of normality are particularly applicable here). Finally, the two additional variables seem intuitively appropriate since MLETTER indicates the relative number of recommendations made to management concerning improvement of the entity's information and control systems and AGGRISKE measures the orientation of management towards risk-taking.

**Explaining Accounts Receivable Errors**

Table 6, Panels A and B, provide the results from estimating the judgment model for AR. The complete (product-moment) correlation matrix is indicated in Panel A. Examination of the correlations for
Table 6. Models for Accounts Receivable Errors

Panel A — Correlation Matrix for Judgment Model

<table>
<thead>
<tr>
<th></th>
<th>AR_{t-1}</th>
<th>EARNPROE</th>
<th>EARNCOMP</th>
<th>CINFOSYS</th>
<th>ATPRESS</th>
<th>AKNOW</th>
<th>GENCONT</th>
<th>ARINCONT</th>
<th>ARJUDG</th>
<th>ARCOMP</th>
<th>ARMATER</th>
<th>GENCONT</th>
<th>ININCONT</th>
<th>INJUDG</th>
<th>INCOMP</th>
<th>INMATER</th>
<th>AKNOW</th>
<th>GENCONT</th>
<th>ININCONT</th>
<th>INJUDG</th>
<th>INCOMP</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR_{t-1}</td>
<td>.38052</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EARNPROE</td>
<td>.00623</td>
<td>.05228</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EARNCOMP</td>
<td>.10132</td>
<td>.01656</td>
<td>.08827</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CINFOSYS</td>
<td>.17230</td>
<td>.05408</td>
<td>.11329</td>
<td>-.02985</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATPRESS</td>
<td>.15234</td>
<td>.02582</td>
<td>.53522</td>
<td>.10422</td>
<td>.23670</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AKNOW</td>
<td>.03120</td>
<td>-.08590</td>
<td>-.40538</td>
<td>-.24211</td>
<td>.02521</td>
<td>-.29565</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GENCONT</td>
<td>-.10531</td>
<td>-.20112</td>
<td>-.14663</td>
<td>.00407</td>
<td>.02499</td>
<td>-.08457</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARINCONT</td>
<td>.06544</td>
<td>-.19443</td>
<td>-.10120</td>
<td>.08551</td>
<td>.18272</td>
<td>.03763</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARJUDG</td>
<td>-.01135</td>
<td>.07659</td>
<td>-.02436</td>
<td>.07869</td>
<td>-.01341</td>
<td>-.02547</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARCOMP</td>
<td>.05803</td>
<td>-.00929</td>
<td>-.05047</td>
<td>.18436</td>
<td>.09313</td>
<td>.03692</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARMATER</td>
<td>-.15826</td>
<td>-.36589</td>
<td>.00010</td>
<td>.11559</td>
<td>-.03142</td>
<td>.05328</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR</td>
<td>.38052</td>
<td>.05228</td>
<td>.08827</td>
<td>.23670</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EARNPROE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EARNCOMP</td>
<td>.10132</td>
<td>.01656</td>
<td>.08827</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CINFOSYS</td>
<td>.17230</td>
<td>.05408</td>
<td>.11329</td>
<td>-.02985</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATPRESS</td>
<td>.15234</td>
<td>.02582</td>
<td>.53522</td>
<td>.10422</td>
<td>.23670</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AKNOW</td>
<td>.03120</td>
<td>-.08590</td>
<td>-.40538</td>
<td>-.24211</td>
<td>.02521</td>
<td>-.29565</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GENCONT</td>
<td>-.10531</td>
<td>-.20112</td>
<td>-.14663</td>
<td>.00407</td>
<td>.02499</td>
<td>-.08457</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARINCONT</td>
<td>.06544</td>
<td>-.19443</td>
<td>-.10120</td>
<td>.08551</td>
<td>.18272</td>
<td>.03763</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARJUDG</td>
<td>-.01135</td>
<td>.07659</td>
<td>-.02436</td>
<td>.07869</td>
<td>-.01341</td>
<td>-.02547</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARCOMP</td>
<td>.05803</td>
<td>-.00929</td>
<td>-.05047</td>
<td>.18436</td>
<td>.09313</td>
<td>.03692</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARMATER</td>
<td>-.15826</td>
<td>-.36589</td>
<td>.00010</td>
<td>.11559</td>
<td>-.03142</td>
<td>.05328</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Panel B — Test of Judgment Model

<table>
<thead>
<tr>
<th>Variable</th>
<th>ANOVA</th>
<th>DF</th>
<th>Sum Squares</th>
<th>Mean Sq.</th>
<th>F</th>
<th>Sig.</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean AR,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Std. Dev (AR,)</td>
<td>= 24.83117</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple R</td>
<td>= .4255</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R Square</td>
<td>= .1810</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Std Dev</td>
<td>= 38.9564</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adj R Square</td>
<td>= .1701</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regression</td>
<td></td>
<td>1</td>
<td>25158.571</td>
<td>25158.471</td>
<td>16.578</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td></td>
<td>75</td>
<td>113820.334</td>
<td>1517.604</td>
<td>SIG .000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>S.E. B</td>
<td></td>
<td>F</td>
<td></td>
<td>Sig.</td>
<td></td>
</tr>
<tr>
<td>AR_{t-1}</td>
<td></td>
<td>.318</td>
<td></td>
<td>.078</td>
<td></td>
<td>16.578</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td></td>
<td>19.106</td>
<td></td>
<td>4.657</td>
<td></td>
<td>16.833</td>
<td></td>
</tr>
</tbody>
</table>
Table 6. Continued

Panel C — Final AR Error Model

<table>
<thead>
<tr>
<th>Variable</th>
<th>$B$</th>
<th>S.E. $B$</th>
<th>$F$</th>
<th>Sig</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>$AR_{t-1}$</td>
<td>.322</td>
<td>.071</td>
<td>20.547</td>
<td>.000</td>
<td>.43071</td>
</tr>
<tr>
<td>MANTURNO</td>
<td>6.648</td>
<td>3.522</td>
<td>3.564</td>
<td>.063</td>
<td>.19591</td>
</tr>
<tr>
<td>AUTOMAT</td>
<td>5.715</td>
<td>2.343</td>
<td>5.950</td>
<td>.017</td>
<td>.23864</td>
</tr>
<tr>
<td>ACONSC</td>
<td>9.712</td>
<td>4.411</td>
<td>4.847</td>
<td>.031</td>
<td>.23127</td>
</tr>
<tr>
<td>Constant</td>
<td>-42.158</td>
<td>15.530</td>
<td>7.369</td>
<td>.008</td>
<td></td>
</tr>
</tbody>
</table>

Mean AR, = 24.83117
Std. Dev. (AR,) = 42.76293
Multiple $R$ = .5929
$R$ Square = .3515
Std Dev = 35.3793
Adj $R$ Square = .3155

ANOVA

<table>
<thead>
<tr>
<th>Regression</th>
<th>Sum Squares</th>
<th>Mean Sq.</th>
<th>$F$</th>
<th>SIG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residual</td>
<td>48856.835</td>
<td>12214.209</td>
<td>9.758</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>90121.970</td>
<td>1251.694</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
the variables and AR, indicates that the only variable with a strong correlation is AR_{-1}. Additionally, with an $F$-statistic of 1.80 required for inclusion in the model, the only variable other than AR_{-1} to enter the regression was CINFOSYS, which is an indicator of the degree of change in the information system during the year.

The initial test of the judgment model included 71 observations, but provided for the exclusion of a company if the company had a missing value on any of the variables in Panel A; when the model was reestimated, including just AR_{-1} and CINFOSYS as estimators (using 77 observations), CINFOSYS did not enter the regression model (nor did any of the other variables) — see Table 6, Panel B. The only variable significant in explaining AR, was AR_{-1}, the absolute value of the error (if any) in the previous year.

Examination of the error/variable correlations for accounts receivable (Table 2, Panel B) indicates that, similar to the inventory model, the following categories contain variables that might improve the initial model presented in Panel B: (1) Category A, the quality of, and situational characteristics for, accounting personnel (AINTEG, ACONSC, and ATURN); (2) Category B, the nature and quality of company's information systems (AUTOMAT); and (3) Category D, the competence, aggressiveness and experience of executive, senior and middle management (MANTURNO).

These five variables plus AR_{-1} were provided as possible explanatory variables for AR, in a stepwise regression model. The results are indicated in Panel C, which reveals substantial improvement over the initial model in Panel B. The adjusted $R^2$ increased from 0.1701 to 0.3155 and the model is statistically significant ($F=9.758, p<.001$) as is each of the five variables ($p<.07$).

As was the case for the inventory errors, the quality of the company's accounting staff has a strong impact on the amount of error, given that ACONSC again is highly significant. MANTURNO and AUTOMAT are also significant in the revised AR model. MANTURNO measures the relative degree of high-level management turnover, with higher scores indicating more turnover and, presumably, more disorientation and stress within the firm. The positive sign of this variable in the model is consistent with this interpretation. Higher scores on the AUTOMAT variable indicate increasing automated financial information systems. It is not clear, however, why a positive relationship should be observed for increasing relative automation and the occurrence of errors.
SUMMARY, LIMITATIONS, AND FUTURE RESEARCH

A summary of the error explanation models developed herein is provided as Table 7. The results suggest that a moderate degree of explanation of errors in accounts receivable and inventory can be achieved given information on the magnitude of the error (if any) in the previous period and information on a few other basic characteristics of the company, in particular, the quality of the entity's accounting staff and measures of aspects of the quality of management. These models seem consistent with intuitive ideas of what aspects of the firm may, directly or indirectly, result in errors in accounts receivable and inventory.

The regression results are based on values of the squared error loss function in which the relative costs of understatement and overstatement of an account balance have an equal impact. Although one might suggest that auditors are more concerned with overstatement than understatement of assets, the equal emphasis seems reasonable when the multiperiod effects of understatement and overstatement of either of these accounts are considered.

Limitations of our statistical analysis should be kept in mind when the results are interpreted. First, results pertaining to levels of statistical significance must be viewed as approximations given the clear nonnormality of the error distributions. Transformations of the errors did not result in significant improvements toward normality of the transformed data. Perhaps an alternative statistical modeling technique to ordinary linear regression (for example, generalized least squares or a nonparametric method) should have been considered. Second, the predictors were measured on a 1 to 7 integer scale and, therefore, the assumption (in a nonfixed effects model) of continuous predictors is violated. Third, cross-validation of the models is desirable via multiple holdout samples within the data or, more importantly, testing the degree to which the models and model relationships generalize to data of a different time period.

Concerning the PMM Risk Database, it is important to note that the manager or in-charge senior who actually completed the questionnaire was also aware of the degree to which errors were detected in the accounts. While many of the questions were designed to measure "objective" aspects of the companies, the possibility of bias, due to their knowledge of the errors, exists for the responses in the database.

A related consideration is that perhaps someone else in addition to the in-charge senior or manager should have been consulted in
Table 7. Summary of Final Inventory and Accounts Receivable Models

Panel A — Inventory (INV.) Model

**Variables in the model:**
1. INV.\(_{-1}\) = Error information from the previous year.
2. ACONSC = Rate the accounting personnel's attitude (Controller and staff). (1=Extremely conscientious, 7=unconscientious)
3. INTRANS = Rate the relative number of unusual transactions (including related party transactions) included in the account. (1=Very few or no unusual transactions, 7=A significant number of unusual transactions)
4. MLETTER = Rate the number and significance of management letter recommendations made to the entity. (1=No recommendations, 7=Many significant recommendations)
5. AGGRISKE = Rate (Executive) management's (i.e., CEO, CFO and other operating officers) aggressiveness in committing the entity to high risk ventures or projects. (1=Extremely conservative, 7=Extremely aggressive)

<table>
<thead>
<tr>
<th>Model</th>
<th>ANOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DF</td>
</tr>
<tr>
<td>Multiple R</td>
<td>.6613</td>
</tr>
<tr>
<td>R Square</td>
<td>.4373</td>
</tr>
<tr>
<td>Std Dev</td>
<td>58.1855</td>
</tr>
<tr>
<td>Adj R Square</td>
<td>.3941</td>
</tr>
<tr>
<td>Variable</td>
<td>B</td>
</tr>
<tr>
<td>INV.(_{-1})</td>
<td>.224</td>
</tr>
<tr>
<td>INTRANS</td>
<td>7.553</td>
</tr>
<tr>
<td>ACONSC</td>
<td>11.152</td>
</tr>
<tr>
<td>MLETTER</td>
<td>5.764</td>
</tr>
<tr>
<td>AGGRISKE</td>
<td>5.738</td>
</tr>
<tr>
<td>Constant</td>
<td>-60.412</td>
</tr>
</tbody>
</table>
Table 7. Continued

Panel B — Accounts Receivable (AR) Model

Variables in the model:
1. AR\(_{t-1}\) = Error information from the previous year.
2. (ACONSC) Rate the accounting personnel’s attitude (Controller and staff). (1=Extremely conscientious, 7=unconscientious)
3. (MANTURNO) Evaluate the extent of management turnover (i.e., CEO, CFO and other operating officers). (1=No turnover, 7=Extremely high turnover)
4. (AUTOMAT) To what degree is (are) the entity’s financial information systems automated. (1=Completely manual, 7=Completely automated)

<table>
<thead>
<tr>
<th>Model</th>
<th>ANOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DF</td>
</tr>
<tr>
<td>Multiple R</td>
<td>.5929</td>
</tr>
<tr>
<td>R Square</td>
<td>.3515</td>
</tr>
<tr>
<td>Std Dev</td>
<td>35.3793</td>
</tr>
<tr>
<td>Adj R Square</td>
<td>.3155</td>
</tr>
<tr>
<td>Variable</td>
<td>B</td>
</tr>
<tr>
<td>INV(_{t-1})</td>
<td>.322</td>
</tr>
<tr>
<td>MANTURNO</td>
<td>6.648</td>
</tr>
<tr>
<td>AUTOMAT</td>
<td>5.715</td>
</tr>
<tr>
<td>ACONSC</td>
<td>9.712</td>
</tr>
<tr>
<td>Constant</td>
<td>-42.158</td>
</tr>
</tbody>
</table>
assessing the general characteristics of the firm. It may have been the case that the people completing the instrument did not have sufficiently complete, reliable information on some of the general entity characteristics.

Finally, a large number of variables at the account level are included in the PMM Risk Database (see the Appendix). Future research could concentrate on more detailed measurements of aspects of companies to investigate further the potential explanatory power of such variables. Several conceptually different kinds of errors (for example, transaction processing vs. asset valuation errors) are possible and specific company characteristics may pertain more directly to some kinds of errors than others.

APPENDIX. CLASSIFICATION OF THE RISK VARIABLES IN THE PMM RISK DATABASE

Panel A — General Variables

A. Quality of, and Situational Characteristics for, Accounting Personnel

12. Rate the level of integrity of the entity's accounting personnel (i.e., Controller and staff) (AINTEG). (1=Extremely high level of integrity, 7=Extremely low level of integrity)

13. Rate the level of knowledge of the entity's accounting personnel (i.e., Controller and staff) in terms of their awareness and understanding of accounting principles and practices and how to apply them. Consider college degrees held, training courses attended, and your observations of the personnel (AKNOW). (1=Extremely high knowledge level, 7=Extremely low knowledge level)

14. Rate the level of experience of the entity's accounting personnel (i.e., Controller and staff). Consider length of experience in current job assignments and all other accounting experience (AEEXPER). (1=Extremely experienced, 7=No experience)

15. Rate the accounting personnel's attitude (Controller and staff) (ACONSC). (1=Extremely conscientious, 7=Unconscientious)

16. Indicate the degree of time pressure on the entity's accounting personnel (i.e., Controller and staff) (ATPRESS). (1=No pressure, 7=Extreme pressure)

17. Evaluate the extent of turnover of the entity's accounting personnel (i.e., Controller and staff) (ATURN). (1=No turnover, 7=Extremely high turnover)

B. Nature and Quality of Company's Information Systems

18. To what degree did the entity's financial information system(s) change in the last fiscal year (CINFOSYS). (1=No change, 7=Significantly changed)

19. To what degree is (are) the entity's financial information systems automated (AUTOMAT). (1=Completely manual, 7=Completely automated)
21. Rate the number and significance of management letter recommendations made to the entity (MLLETTER). (1=No recommendations, 7=Many significant recommendations)

22. Indicate management's response to management letter recommendations (MLRESPON). (1=Recommendations fully implemented, 7=No recommendations implemented)

C. General Controls

20. Rate the overall level of general controls, including potential for management override. Consider factors such as organizational structure, documentation policies, existence of budgets and comparison of budgets to actual results, and existence of an internal audit department (GEN-CONT). (1=Extremely high level of general controls, 7=No general controls)

D. Competence, Aggressiveness and Experience of Executive, Senior and Middle Management

6. Rate the level of competence of the entity's management (i.e., CEO CFO, and other operating officers)
   a. Executive management (MCOMPETE). (1=Extremely competent, 7=Extremely incompetent)
   b. Senior/middle management (MCOMPTEM). (1=Extremely competent, 7=Extremely incompetent)

7. Rate management's (i.e., CEO, CFO and other operating officers) aggressiveness in committing the entity to high risk ventures or projects.
   a. Executive management (AGGRISKE). (1=Extremely conservative, 7=Extremely aggressive)
   b. Senior/middle management (AGGRISKM). (1=Extremely conservative, 7=Extremely aggressive)

8. Rate the experience level of the entity's management (i.e., CEO, CFO, and other operating officers).
   a. Executive management (EXPERE). (1=Extremely experienced, 7=Extremely inexperienced)
   b. Senior/middle management (EXPERM). (1=Extremely experienced, 7=Extremely inexperienced)

11. Evaluate the extent of management turnover (i.e., CEO, CFO, and other operating officers) (MANTURNO). (1=No turnover, 7=Extremely high turnover)

E. Degree of Emphasis by Management on Accounting Earnings

9. Rate management's emphasis on meeting earnings projections.
   a. Executive management (EARNPROE). (1=Little emphasis, 7=Significant emphasis)
   b. Senior/middle management (EARNPROM). (1=Little emphasis, 7=Significant emphasis)

10. Indicate the degree to which compensation of the entity's management (i.e., CEO, CFO, and other operating officers) is based on operating results.
    a. Executive management (COMPENE) (1=No relationship to operating results, 7=Totaly related to operating results)
b. Senior/middle management (COMPENM) (1=No relationship to operating results, 7=Totally related to operating results)

23. Evaluate the entity's liquidity position. Consider current ratio, quick ratio, cash flow projections and similar indicators of liquidity (LIQUID). (1=Extremely liquid, 7=Extremely illiquid)

24. Evaluate the degree to which the entity's liquidity position changed from that of the prior period (LIQUIDG). (1=Significantly more liquid, 7=Significantly less liquid)

25. If the entity has debt with restrictive loan covenants, rate the ease with which the entity complied with the restrictive covenants (COMPLOAN). (1= Easily complied with all restrictive covenants, 7=Not in compliance with any restrictive covenants)

26. Evaluate the entity's net earnings relative to its size (i.e., total assets, total equity) (RETURNA). (1=Excellent return, 7=Negative return)

27. Evaluate the degree to which the entity's results of operations are broad-based (RESOTHER). (1=Extremely broad-based, 7=Extremely dependent on a limited number of segments or products)

28. Evaluate the entity's net earnings relative to other entities in the same industry (RELEARN). (1=Significantly higher, 7=Significantly lower)

29. Evaluate the entity's most recent net earnings relative to the entity's net earnings in the prior period (EARNCOMP). (1=Significantly higher, 7=Significantly lower)

30. Rate the profitability of the entity's industry relative to other industries (INDPROF). (1=High level of profits, 7=High level of losses)

F. Nature of Firm's Operating Environment

1. Evaluate the degree to which the entity's operations are centralized (CENTOPER). (1=Highly centralized, 7=Highly decentralized)

2. Evaluate the number of "outsiders" on the board of directors (OUTSIDER). (1=High number of outsiders, 7=No outsiders)

3. Evaluate the rate of turnover of "outsiders" on the board of directors (OUTSTURN). (1=No turnover, 7=Extremely high turnover)

4. Evaluate the frequency of recent allegations and/or actions brought against either management or the company by regulatory bodies (ALEFGRE). (1=Very frequent, 7=Very infrequent)

5. Evaluate the seriousness of recent allegations and/or actions brought against either management or the company by regulatory bodies (ALEFGSER). (1=Extremely serious, 7=Very trivial)

29. Evaluate the overall business outlook for the entity's industry (BUSOUT). (1=Extremely healthy, 7=Extremely unhealthy)

31. How sophisticated is the level of technology employed to produce the entity's major products or service lines (TECHPRO). (1=Extremely unsophisticated, 7=Extremely sophisticated)

32. How sophisticated is the technology of the entity's major products or service lines (TECHPSER). (1=Extremely unsophisticated, 7=Extremely sophisticated)

33. Evaluate the potential for the entity having a continuing interest in the product or service after the point of sale (e.g., buy-back, warranties) (CONTPROI). (1=Little or no potential for continuing involvement, 7=High degree of continuing involvement)
34. How competitive is the entity's industry (INDCOMP). (1=No other competitors, 7=Extremely competitive)
35. Evaluate the degree and complexity of local, state, and federal regulations affecting the entity's industry (REGCOMP). (1=Substantial regulation, 7=Little regulation)
36. Evaluate the rate of change in the entity's industry. Consider changes such as the entry or exit of competitors, introduction of new products and technology, changes in leadership positions among companies in the industry, and rate of growth (INDCHG). (1=No change, 7=High level of losses)
38. Evaluate the effect on the entity's business of fluctuations in interest rates (INTRATE). (1=No effect, 7=Significant effect)
39. Evaluate the effect of inflation on the entity's business, including the entity's ability to pass on inflationary increases to customers (INFLAT). (1=No effect, 7=Significant effect)
40. Evaluate the effect of fluctuations in foreign exchange rates on the entity's business (EXCHGR). (1=No effect, 7=Significant effect)

G. Legal and Political Environment
41. Evaluate the effect of recent changes in the political environment on the entity's business. Consider the effect of new laws and regulations, nationalization of foreign investments, trade barriers, etc. (POLENVE). (1=No effect, 7=Significant effect)

Panel B — Account Specific Variables
1. Rate the degree to which judgment (including estimates) was required in arriving at the entries to the account. (*JUDG). (1=Little or no judgment was required, 7=Extreme judgment was required)
2. Rate the degree of complexity underlying the entries made to the account. For example, summarizing cash receipts usually is not complex, whereas calculating the liability and expense related to federal income taxes is often highly complex. (*COMP). (1=Little or no complexity underlying the entries, 7=An extremely high degree of complexity underlying the entries)
3. Rate the effectiveness of the internal controls affecting the account. (*INCONT). (1=Effective internal controls, 7=Little or no internal control)
4. Rate the susceptibility to theft or fraud of the asset, liability, or equity underlying the account. (*FRAUD). (1=Little or no susceptibility to theft or fraud, 7=An extremely high degree of susceptibility)
5. Rate the materiality of the account balance (*MATER). (1=A highly immaterial account, 7=A highly material account)
6. Evaluate the make-up of the account by rating the relative size of its individual components (*FSIZE). (1=An account with many individually immaterial items, 7=An account with a small number of material items)
7. Rate the relative number of unusual transactions (including related party transactions) included in the account. (*TRANS). (1=Very few or no unusual transactions, 7=A significant number of unusual transactions)
8. Rate the degree to which year-to-year characteristics of the account showed unusual relationships (e.g., relative size of discrepancies between budgeted and actual amounts or unexpected year-to-year fluctuations in the balance or in transaction volume). (*RELAT). (1=No unusual relationships or trends, 7=Several unusual relationships or trends)

9. Rate the seasonality of the activity represented by recordings in the account. (*SEAS). (1=Little or no seasonality, 7=Significant seasonality)

* Indicates that the AR----- or IN----- labeling applies for the accounts receivable and inventory accounts, respectively.

NOTES

*This research was started while the second author was an Audit Research Fellow in the Audit Research Group at the Executive Office of Peat, Marwick, Mitchell & Co. We acknowledge the important contributions of Joe Diebold, Robert Henarie, and Doug Wacek. Participants in the Accounting Workshop at the University of Minnesota and the Accounting Research Forum at Boston University provided useful comments on this research. Discussions with Andrew Bailey and Shyam Sunder, and the comments of the reviewer, were particularly helpful.

1. The accounts receivable and inventory accounts were selected due to their importance for both the balance sheet and income statement. Also this part of the database was the only segment ready for usage at the time the study was conducted.

2. As a practical matter, Coopers & Lybrand (1983) consider the inherent and control risks simultaneously, partitioning errors according to errors subject to the internal controls vs. situational risk factors.


4. Development of the PMM Risk Database proceeded as follows. A random sample of 1,209 clients was selected from the firm's master list of clients. Situations that did not qualify for inclusion in the sample (nonaudit engagements or new clients therefore not having prior year audit information) were excluded leaving 873 firms. A comprehensive questionnaire was mailed to the partner responsible for the audit engagement. The questionnaire was completed by the engagement manager or in-charge senior responsible for the audit. Responses were not received for 176 firms, and 170 responses were unusable since they were incomplete or the company had a different operating environment from a classification of 10 major industries, resulting in 527 usable client descriptions in the database. In this study, we analyze the 82 responses from the manufacturing industry (the segment of the database that is available for analysis at this time).

5. The partners (and therefore the managers or in-charge seniors who provided the data) were told to exclude audit differences that were due to expected, recurring year-end bookkeeping adjustments (for example, cash to accrual accounting adjustments, accrual of income taxes, determination of LIFO reserves and allocation of variances from standard costs to inventories). The coders of the data checked the responses to make sure expected adjusting entries were not included in the error distributions.

6. The measure of client size used here ("audit gauge") is a function of either the magnitude of revenues or total assets of the firm at the current
point in time. The actual function is gauge = 1.6 size², where size is the larger of revenues or assets for the year.

7. Festge (1979, p. 105) reports that overstatements are common for AR while it is common to have both understatements and overstatements of INV amounts.

8. Johnson, Leitch, and Neter (1981, pp. 283ff) reported a similar finding.

REFERENCES


Estimation of Errors in Financial Statements: Discussion

The Willingham/Wright paper examines financial statement errors discovered in year-end audits and attempts to determine their estimability given specific knowledge of a firm. Because of the almost total lack of research based upon "real" audit data, this paper is a welcome addition to the auditing literature. However, in reviewing the paper, I have four major concerns: (1) the lack of a conceptual framework for structuring the study; (2) the integrity and validity of the data base; (3) the overall approach to analyzing the data; and (4) the possible implications of the research for practice.

LACK OF CONCEPTUAL FRAMEWORK

The paper begins with a general reference to the audit risk model set forth in Statement on Auditing Standard (SAS) No. 47 (AICPA, 1983). It is the authors' apparent intention to use this model as a basis for structuring the study and analyzing the data. However, there is little discussion of how the SAS No. 47 model might be operationalized in practice and, as the paper progresses, the auditors do not focus
upon the model or its components. Rather, the study focuses on the overall (aggregate) estimation of financial statement errors based upon a variety of factors. The authors do not attempt to relate these factors directly to the audit risk model or its components. Consequently, the results of the study cannot be used to infer (in a descriptive sense) reliance factors to the audit risk model or its components (inherent risk, control risk, or detection risk).

INTEGRITY AND VALIDITY OF THE DATA BASE

The paper is very brief in its description of the data base and how it was generated. Apparently, the data base (the Peat, Marwick, Mitchell & Co. Risk Data Base) was developed from randomly selecting 1,209 clients from a master list of Peat, Marwick, Mitchell's clients. However, the selection process is not described. In addition, bias could have been introduced because only 873 clients (out of the 1,209) were actually surveyed, 176 nonresponses occurred, and 170 nonuseable responses were received. The study does not contain any analysis of potential nonresponse or nonuseable response bias nor does the paper express any concern over the potential existence of bias.

The research instrument was not provided in the paper. As a minimum, a copy of the instrument should have been included (perhaps as an appendix) so that the reader could interpret the reasonableness of the survey questions, and so on (only excerpts from the research instrument are provided in the Appendix).

The development of the research instrument (questionnaire) is not described. Therefore, it is not known whether the questionnaire was pretested. In reading excerpts from the questionnaire I noted several questions for which the auditors could have experienced difficulty in providing accurate, meaningful answers. For example, Questions 12 and 13 request the auditor's assessment (in aggregate) of the integrity and knowledge of the entity's accounting personnel. In most audit situations, the audit engagement personnel assess the integrity and competency of individual personnel, such as the controller or the assistant controller. These individual assessments are then used in structuring the audit if the assessment affects the likelihood of material error in individual account balances or classes of transactions. Usually, assessments are not developed in an aggregate sense for the reason that such assessments are difficult to make and even more difficult to interpret. For example, if a controller is assessed as incompetent, but the assistant controller is assessed as competent, how should the
auditor assess the competency of accounting personnel in an aggregate sense? Moreover, what effect might such an assessment have on the audit? For this reason, the answers to these questions may contain significant noise.

A similar question requires the auditor to assess the effectiveness of internal controls over a particular account (accounts receivable and inventory in this paper). For the reasons already mentioned, answers to these questions may contain noise. In many situations audit personnel would have difficulty rating the overall effectiveness of internal controls over an account or class of transactions (in aggregate) as opposed to rating the effectiveness of controls over individual transaction functions, such as the shipping function or billing function.

There also is potential noise created in the data base because three levels of engagement personnel could have completed the questionnaire (that is, the partner, the manager, or the senior). Apparently, there was no control over which individual actually did complete the questionnaire. This lack of control also reduces the internal validity of the data base and is especially important given that many of the questions required subjective assessments by the auditor respondent.

The paper does not describe any attempt at cross-validation of the answers to the questionnaire. For example, several questions ask the auditors to rate the liquidity of the entity under audit, but apparently no attempt was made to cross-reference the answers to an actual assessment of liquidity vis-à-vis financial statements of the client. Another example of a possible cross-validation would be a comparison of the answers to the internal control effectiveness questions with whether the auditors planned reliance on the control system to reduce their substantive testing.

Additionally, the data base (and its analysis) is suspect because of the apparent existence of several outliers which could bias the data. For example, the mean accounts receivable error for period t is negative while the mean accounts receivable error for period (t-1) is positive. Also the mean, median, and mode statistics differ significantly which suggests the possible existence of outliers. As a minimum, more descriptive statistics for the data base (including client financial statement data) should have been provided.

The data were collected in late 1982 and were based upon audits for the period of 1980–81. Thus, the data reflect pre-SAS No. 39 and No. 47 audits. It therefore seems inconsistent to use SAS No. 39 and No. 47 concepts as a basis for developing, justifying, and structuring the study.
Finally, a potential for bias exists in the database because the data were gathered only on Peat, Marwick, Mitchell & Co. audit clients and only by Peat, Marwick, Mitchell & Co. personnel. Thus, the answers provided could be conditioned by firm training materials, firm audit manuals, and firm client selection criteria. However, this concern must be balanced against the difficulty of obtaining real audit data. Appropriate caveats, nevertheless, should have been added with respect to these potential biases.

OVERALL APPROACH TO ANALYZING THE DATA

The logic of the approach the authors used to analyze the data is not clearly set forth. Specifically, the authors initially asked two experienced auditors to develop a judgmental model of financial statement errors. This judgmental model was developed by asking the auditors to list ten variables associated with financial statement errors from a preconceived list which was developed by the authors. The fact that a preconceived list was used adds potential bias to the judgmental model. In addition, it is not clear from reading the paper which specific items were included by the auditors in their judgmental model. For example, did auditors include (or were they allowed to include) internal accounting controls, general controls, or last year's error as factors which would be useful in estimating current year's errors?

The judgmental models were subject to empirical testing vis-à-vis stepwise regression but generally did not perform well. After the judgmental models had been run and performed poorly, the authors extended the analysis to include other potential models. However, it is not clear how these other potential models were developed. For example, the authors ran correlations of variables associated with financial statements errors against the actual reported errors. However, in developing alternative explanatory models some factors with high correlations were used while others were not.

In the analysis of the data, the authors did not segregate types of errors. Consequently, computational errors, classification errors, judgmental errors, and so on, are analyzed in aggregate. Because the factors generating different types of errors (for example, classification errors versus computational errors) are likely to differ, any model based upon an aggregation of errors would, a priori, explain poorly. This was confirmed by the results of the research. I suggest the authors consider partitioning the data base by error type and reanalyzing the data.
In analyzing the data, the authors used the absolute value of the net dollar error divided by gauge as the dependent variable. This choice potentially adds bias because not all audit firms adhere to the audit gauge philosophy, which implies that materiality (and consequently the extent of testing) is greater (in a relative sense) for smaller clients than for larger clients. In addition, because absolute values were used, the authors treated overstatement and understatement errors as equally important. However, such equal treatment does not conform to practice since, for example, auditors often search for overstatement errors in assets more extensively than for understatement errors.

The effects of distributional characteristics of the data on the estimation models was not adequately addressed in the paper. The authors do indicate that a revision of the paper is forthcoming in which various transformations are utilized to reduce the effects of the distributional characteristics of the data. I chose not to comment upon these possible effects because of the forthcoming revisions which I presume will address these issues.

IMPLICATIONS FOR AUDITORS

The implications for auditors, given the models developed in the paper, are not clear. It would have been useful, for example, if the authors had provided some illustrations of how the practitioner might employ the models actually derived in the study.

SUMMARY

This paper has some very interesting ideas and is a good starting point to explore further analyses of financial statement error estimation. Overall, the analysis of the data is too broad, however, and I recommend that the authors consider partitioning the data base and reanalyzing it (for example, by error type).

Several minor concerns which I found annoying included the difficulty of following the tables and the notation utilized for identifying the explanatory factors. Finally, I noted that in the initial stages of the paper, the authors define audit risk, inherent risk, control risk, and detection risk, but that these definitions are inconsistent with SAS No. 47. Perhaps the authors intend the definitions used in the paper to be superior to the definitions in SAS No. 47. If this were the case, the authors should have attempted to justify the definitions used.
REFERENCES
Estimation of Errors in Financial Statements: Discussion

The Willingham and Wright paper is a significant contribution to information on the causes of errors in financial statements. The firm of Peat, Marwick, Mitchell & Co. is to be complimented for making the results of this study available, particularly in consideration of the cost of gathering the required data. This type of research should be pursued, whether or not the present study eventually is able to identify the causes of errors in financial statements, so that extant, hypothetical, models for setting audit scope can be developed into useful tools for application in practice.

It should be understood from the beginning that these comments are prepared from my perspective as a practitioner with an interest in applying the results of audit research to practice. As Willingham and Wright make clear, this specific research is not complete and, doubtless, additional refinements will be attempted before completion of this project.
THE AUDIT RISK MODEL

The audit risk model proposed in *Statement on Auditing Standards (SAS) No. 39* (AICPA 1981) and further discussed in *SAS No. 47* (AICPA 1983), implies that the task of setting audit scope involves controlling the risk of material misstatement of financial statements through assessments of (1) the likelihood that such errors exist undetected by the client's system of internal control and (2) the effectiveness of the auditing procedures which will be used to detect such errors. There are a number of incentives to develop a practical application for this hypothetical risk model.

First, an empirically based model would significantly improve the quality of evidence supporting the auditor's scope selection. It might also allow less experienced personnel to perform a greater proportion of audit planning. Finally, such a model should be useful in the auditing standards-setting process, by facilitating the study of the effects of proposed standards prior to adoption.

The Willingham and Wright study deals with estimation of the absolute value of errors. The rationale stated is that the auditor's loss function is assumed to be symmetric, that is, that the auditor is indifferent to the direction of error since the consequences of equally sized errors in either direction are the same. The use of this presumption significantly reduces the potential benefit of the results of a statistically successful model of this type. Neither the auditor's loss function nor, as important, the company's loss function are symmetric. Litigation and decreases in share value tend to be more associated with overstatements of assets and understatements of liabilities, and their corresponding effect on reported earnings. Indeed, the accounting principle of conservatism embraces and encourages this bias.

I believe the limitations of the research approach used in this study preclude accepting the researchers' conclusion that a moderate degree of estimation of errors in accounts receivable and inventory was demonstrated. However, neither are the results persuasive that such relationships do not exist. The primary value of the paper is, first, in presenting initial conclusions regarding the likelihood of financial statement errors and, then, in providing a framework and a basis for work of this type in the future. Some of these improvements and refinements will, no doubt, come as a result of further work with the Peat, Marwick, Mitchell and Co. data base. Others will require the creation of new data bases containing additional or revised information.
METHODOLOGY

There are certain incomplete areas of the paper, acknowledged by the authors, which raise significant questions about the overall effectiveness of the research. One such question is whether or not certain types of data transformations might improve the results of the model. In addition, the possible effects of the assumption that overstatement and understatement errors are of equal weight should be evaluated, as should the effect of the engagement teams' prior knowledge of error rates on their responses. A partial response would be to attempt to use quantifiable or measurable proxies for specific key factors to bring more objectivity to the measurement process.

The database contained data on errors identified in financial statements as a result of audit procedures. Consequently, the error rates encountered represent those which were undetected and uncorrected by the system of internal control and which were subsequently detected by the year-end audit process. Thus, this study primarily deals with the combined effect of inherent risk and control risk, assuming that audit detection risk is equal to zero. A linear scale (one to seven) was used to quantify the 55 independent variables which were included in the data base. The paper would have been helped by the provision of scattergrams showing the dispersion of responses by attribute. Apparently this will be a subject of further work on the possible effect of transformations on the results.

A review of the narrative description of the 55 independent variables used in the data base reveals the inclusion of all of the general factors which auditors would identify as having an effect on the occurrence rate of errors. However, it is my opinion that many of these variables are intercorrelated. For example, an assessment of the general level of controls (GENCONT), by definition, includes an assessment of the level of controls over specific accounts (ININCONT or ARINCONT), as well as assessments of the quality of the entity's systems (CINFOSYS, AUTOMAT, MLETTER, MLRESPON), and the quality of its accounting personnel (AINTEG, AKNOW, AEXPER, and so on).

The final form of both models includes the actual prior period error rate. Since many of the variables which impact error rates may be expected to change slowly over time, it is reasonable to expect that the prior period error rate may duplicate the information content of many of the possible current period explanatory variables. Consequently, inclusion of the prior period error rate in the step-wise regression may have excluded other variables whose informational
content is included in the prior period error rate. Techniques such as factor analysis, therefore, might be used to improve the model. In any event, I believe the problem of multicollinearity is pervasive in this study and should be addressed. The possible explanatory variables were quantified on a linear scale of one to seven, based on the engagement team’s assessment. I believe that several of these variables could be measured more objectively by other means. Specifically, a number of the variables relate to the financial condition of the company both over time and against its peers. Rather than using the engagement team’s subjective rankings, certain financial statement measures of liquidity, financial condition, and earnings might have been used.

ERROR DEFINITION

An error was defined as the difference between the audited and book value of inventory and receivables. Consequently, adjustments affecting income and classification errors were not distinguished. Although both types of errors affect the reported balances, there may be differences in the processes and resulting relationships which generate different types of misstatements. Later studies should differentiate error types to permit study of whether the underlying error generating processes differ.

Certain of the explanatory variables may be much more correlated with errors in a specific direction or of a specific type. For example, it would seem that various conditions which might imply a management bias toward overstatement of earnings (for example, management bonus plans based on accounting earnings, substandard earnings, and so on) would tend to produce overstatements of income through management modification of the results of transaction processing. Deficiencies in the general quality of transaction controls and accounting personnel, however, might produce a random pattern of error direction and size.

STATISTICAL RESULTS

Although moderately acceptable statistical results were obtained from the final forms of the models, these results failed to confirm the judgmental models elicited from experienced auditors. In particular, the auditor’s judgment that the quality of general and account-specific controls, and the degree of emphasis on accounting earnings, would be key factors in estimating error was not supported by the final model. The failure of the judgmental model could result from
the fact that the number of possible explanatory variables which the auditors were allowed to select was limited to ten, and that only two auditors were polled in developing the judgmental model.

CONCLUSION

Demonstration of the existence of one or more error-generating processes and the study of their nature is of critical importance to the development of structured audit scope decision methodologies. Although this study professes to have demonstrated a "moderate" degree of estimation of errors in accounts receivable and inventory, the limitations of its methodology make it difficult to accept this conclusion. The primary value of this work at its present point is to direct attention to the practical problems of researching error occurrences and to stimulate further work on this important issue.

REFERENCES


FIVE
Michael Gibbins and Craig Emby*

University of Alberta

Evidence on the Nature of Professional Judgment in Public Accounting

Public accountants, as professionals, exercise judgment in many circumstances. The importance of professional judgment is recognized by standard-setting bodies; it is frequently mentioned in the accounting and auditing standards of the American Institute of Certified Public Accountants, the Canadian Institute of Chartered Accountants, and the Financial Accounting Standards Board (Gibbins 1983). Research interest in the judgment of physicians, auditors, bankers, bond raters, and other professionals has been substantial and many accounting and auditing judgment studies already have been published.¹

Much has been accomplished in modeling judgment processes that appear to exist in the general population and therefore, carry implications to public accounting. Several experimental studies have examined such implications by using auditors as subjects. However, the applicability of both the general and the auditor-based research findings to the day-to-day circumstances in which public accountants

Editors' Note: This paper was anonymously refereed prior to acceptance.
exercise their judgment is not clear, partly because those circumstances have not been well defined. Because professional judgment is an interaction between the cognitive processes of professionals and the demands of the professional environment, including, for example, expectations about consultation and documentation, the nature of professional judgment itself is not well specified.  

This paper is intended to improve the specification of the nature of professional judgment in public accounting (henceforth, PJPA). PJPA is a complex process, carried out in a complex environment, so its nature will be examined from several directions. The purpose is to display the complexity, to organize the descriptions rather than simplify them, in order to provide information that is rich enough to be meaningful to practitioners and useful to other researchers. In the course of doing this, some hypotheses about the process will be tested.

CONCEPTUAL BACKGROUND

Judgment may be viewed as an action ("I made a judgment") or as a process ("I used my judgment"). In this paper, the latter sense will be used: the word "judgment" and the expression "PJPA" will, with a single exception to be identified later, be restricted to the cognitive process; behaviors such as choosing, acting, or announcing a decision will be considered as results.

A somewhat unconventional questionnaire was used to gather the data. It was based on a cognitive learning model, developed from psychological and accounting/auditing research and elaborated on in Gibbins (1984), that divides cognitive processes generally into four highly interrelated groups: event perception, memory organization and accumulation, response preference generation, and response preference evaluation. Two interrelated factors also are involved in this cognitive learning model: the nature of the environment within which the judgment process operates (including the presence or absence of feedback), and the extent to which learning affects the operation of that process (experienced judgment).

The questionnaire was constructed to provide evidence, collected from people at various levels in public accounting firms in five countries, about the nature of PJPA via the four groups and two factors already described. As the results are set out, the specific purpose of each part of the questionnaire will be explained. Subsequently the conclusions tie the various results together.
THE RESEARCH QUESTIONNAIRE

Respondents were sought a few at a time in each of several public accounting firms in several countries and asked to provide advice to the author, as a professional colleague. The questionnaire was called "Inventory of Professional Judgment Process," and was described to respondents as being intended to gather information systematically from "people holding various positions in several different public accounting firms in several countries," and as being challenging because it dealt with serious issues.

In order to orient the questionnaire to PJPA as experienced by the respondents, an "interview on paper" approach was used: in several sections, respondents were asked to describe experiences in their own words and then were asked a number of questions about those experiences. The intent was to leave the respondent "in control" of the inventory, and thus avoid suggesting or imposing any view of PJPA via the questionnaire. Questions were framed as requests for collegial assistance in understanding PJPA; respondents were promised anonymity for themselves and their firms. The questionnaire was tested and refined, prior to its administration, through a large number of interviews with practicing public accountants, primarily from Canada but also from the United States and Europe.

Two 12-page versions of the questionnaire, called henceforth the "yellow" and "blue" questionnaires, were developed. The first three pages of each were identical, providing an introductory letter from the author, asking some background questions and specifying some terms that would be used. No respondent received both versions. This procedure reduced the time required to answer the questionnaire and allowed us to use some questions in both versions, thus providing a cross-check on the validity of the results.

Respondents were obtained by contacting a number of public accounting firms (in Canada, the United States, the United Kingdom, the Netherlands, and Australia), and asking each to distribute the two questionnaire versions to professionals at four different levels: a partner, a manager, a supervisor (or equivalent title), and an audit assistant (or equivalent title). This selection procedure was used to ensure a wide range of experience in PJPA and a range of cultural and organizational environments.

The first page of each questionnaire was a cover letter and the second page asked about the respondent's position and number of years in public accounting. In accordance with the conceptual back-
ground above, we introduced the following definitions on the third page:

In this questionnaire, "judgment" will be used to refer to the process leading up to a decision or action. The decision or action will thus be said to be on the basis of a person's professional judgment. Consequences or results in turn will be said to follow from the decision or action. Thus,

![Diagram showing the process from professional judgment to decision or action to consequences or results.]

The judgment process may be long or short, conscious or unconscious, broad or narrow. The decision may involve action or inaction. The consequences or results may be immediate or delayed, major or minor, clear or obscure. The questionnaire seeks your help in clarifying some of these matters.

This quote was the only statement made in the questionnaires about PJPA so that the respondents would not be constrained in expressing their views about the subject.

Use of a questionnaire raises the likelihood of reactivity (demand) effects. Demand effects might exist to the extent that respondents determine what the investigator seems to be after, and to the extent that PJPA may involve cognitive processes that are not fully conscious or ones the respondent may not have thought about until asked (Gibbins 1984). The likelihood of reactivity was reduced by leaving topics open and avoiding hints about the researcher's orientation. In addition, the respondent usually was asked to provide a description of a topic first and then was questioned about it on a later page. This procedure should reduce the possibility that the questions would influence the nature of the description, thus enhancing the likelihood that the respondent would feel a commitment to his/her own description. Since retrospection was required by some areas of the questionnaire, questions and analyses also were designed to reduce the retrospective rationalizing that is likely in such circumstances (Gibbins 1984).

**RESPONDENTS**

There were 176 potential respondents to the two forms of the questionnaire of which 136 were returned. Thus the overall response
rate was 77 percent. Response rates to the two questionnaire versions were almost identical: 67 of 88 “yellow” questionnaires were returned (76 percent) and 69 of 88 “blue” ones (78 percent), and response rates were almost the same within each country. Response rate statistics by country follow (a sixth “country,” “Can-US” is shown because one firm sent its U.S. responses via a Canadian office where they were inadvertently mixed with its Canadian responses):

<table>
<thead>
<tr>
<th>Country</th>
<th>Distributed</th>
<th>Response</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canada (4 firms)</td>
<td>32</td>
<td>28</td>
<td>88</td>
</tr>
<tr>
<td>United States (4 firms)</td>
<td>32</td>
<td>28</td>
<td>88</td>
</tr>
<tr>
<td>Can-US (1 firm in each country)</td>
<td>16</td>
<td>16</td>
<td>100</td>
</tr>
<tr>
<td>United Kingdom (3 firms)</td>
<td>24</td>
<td>21</td>
<td>88</td>
</tr>
<tr>
<td>Australia (3 firms)</td>
<td>40</td>
<td>16</td>
<td>40</td>
</tr>
<tr>
<td>Netherlands (4 firms)</td>
<td>32</td>
<td>27</td>
<td>84</td>
</tr>
<tr>
<td>Totals</td>
<td>176</td>
<td>136</td>
<td>77</td>
</tr>
</tbody>
</table>

Table 1 summarizes the respondents’ positions and experience. Panel (A) of the table indicates that proportionately more respondents put themselves in the “supervisors” classification than in the other

Table 1. Position and Experience of Respondents

(A) Position by Country

<table>
<thead>
<tr>
<th>Country</th>
<th>Partners</th>
<th>Managers</th>
<th>Supervisors</th>
<th>Assistants</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canada</td>
<td>5</td>
<td>8</td>
<td>10</td>
<td>5</td>
<td>28</td>
</tr>
<tr>
<td>U.S.A.</td>
<td>5</td>
<td>7</td>
<td>8</td>
<td>8</td>
<td>28</td>
</tr>
<tr>
<td>Can-US</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td>U.K.</td>
<td>6</td>
<td>3</td>
<td>7</td>
<td>5</td>
<td>21</td>
</tr>
<tr>
<td>Australia</td>
<td>3</td>
<td>2</td>
<td>9</td>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td>Netherlands</td>
<td>5</td>
<td>7</td>
<td>7</td>
<td>8</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>31</td>
<td>47</td>
<td>30</td>
<td>136</td>
</tr>
</tbody>
</table>

(B) Average Years of Experience by Country and Position

<table>
<thead>
<tr>
<th>Country</th>
<th>Partners</th>
<th>Managers</th>
<th>Supervisors</th>
<th>Assistants</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canada</td>
<td>16.8</td>
<td>7.6</td>
<td>3.4</td>
<td>1.2</td>
<td>6.6</td>
</tr>
<tr>
<td>U.S.A.</td>
<td>15.8</td>
<td>9.4</td>
<td>4.9</td>
<td>1.3</td>
<td>6.6</td>
</tr>
<tr>
<td>Can-US</td>
<td>17.3</td>
<td>9.0</td>
<td>3.1</td>
<td>.7</td>
<td>7.8</td>
</tr>
<tr>
<td>U.K.</td>
<td>20.2</td>
<td>9.0</td>
<td>4.5</td>
<td>2.0</td>
<td>9.0</td>
</tr>
<tr>
<td>Australia</td>
<td>21.8</td>
<td>13.3</td>
<td>5.9</td>
<td>2.0</td>
<td>9.3</td>
</tr>
<tr>
<td>Netherlands</td>
<td>17.8</td>
<td>12.8</td>
<td>13.6</td>
<td>6.6</td>
<td>12.1</td>
</tr>
<tr>
<td></td>
<td>18.2</td>
<td>9.9</td>
<td>5.8</td>
<td>2.8</td>
<td>8.5</td>
</tr>
</tbody>
</table>
three classifications, especially the Australians. Panel (B) of the table shows that the respondents averaged 8.5 years of experience, but varied within positions as would be expected. (For the respondents to the yellow questionnaire, position correlated .80 with years of experience; for those responding to the blue version, the correlation was .62.) The panel also shows that the non-North American respondents, especially the Dutch, generally were more experienced than the North Americans. The "supervisor" classification again shows an anomaly: this time the Dutch supervisors stand out. Perhaps the meaning of that classification was not clear-cut for the Australians and the Dutch.

QUESTIONNAIRE SECTIONS: EXPLANATIONS AND RESULTS

Characterization of PJPA

Explanation and Design

Most sections of the questionnaire dealt with particular aspects of PJPA. To provide a general background and insight into the way public accountants themselves characterize PJPA, each version of the questionnaire contained a section asking for perceptions about several potential facets of PJPA. The phrases used in the two parts of the questionnaire were taken from remarks about expertise, responsibility, morals, routine, and so on, made by public accountants interviewed during the questionnaire's development and testing.

In the yellow questionnaire, respondents were given a list of 18 phrases and were asked to use them to define PJPA. Following remarks made during the developmental interviews, the phrases were grouped into "possible qualities of the accountant or auditor exercising PJPA" and "possible qualities involving the PJPA situation." Respondents were asked to indicate which of the four categories appropriately relate each phrase to PJPA. The categories ranged from "essential to PJPA" to "irrelevant to PJPA." Space also was provided for respondents to add and categorize additional phrases; however, only a few did. Table 2 contains the 18 phrases and the four category definitions.

In the blue questionnaire, respondents were given 15 statements drawn from remarks made by public accountants in the earlier interviews and were asked to categorize each as being true for all, most, some, or no professional judgments made by the respondent. Table 3 contains the 15 statements and the four category definitions.
Table 2. Characterizations of PJPA (Yellow Questionnaire)

<table>
<thead>
<tr>
<th>Possible Qualities of the Accountant or Auditor Exercising PJPA:</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y1. High expertise</td>
<td>2</td>
<td>4</td>
<td>33</td>
<td>28</td>
</tr>
<tr>
<td>Y2. Analytical ability</td>
<td>0</td>
<td>2</td>
<td>19</td>
<td>46</td>
</tr>
<tr>
<td>Y3. Knowledge of theory</td>
<td>2</td>
<td>10</td>
<td>33</td>
<td>22</td>
</tr>
<tr>
<td>Y4. Ability to implement</td>
<td>6</td>
<td>13</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>Y5. Practical experience</td>
<td>0</td>
<td>5</td>
<td>26</td>
<td>36</td>
</tr>
<tr>
<td>Y6. Nonanalytical intuition</td>
<td>10</td>
<td>28</td>
<td>23</td>
<td>6</td>
</tr>
<tr>
<td>Y7. High ethics/morals</td>
<td>9</td>
<td>14</td>
<td>16</td>
<td>28</td>
</tr>
<tr>
<td>Y8. Prudence</td>
<td>6</td>
<td>19</td>
<td>29</td>
<td>13</td>
</tr>
<tr>
<td>Y9. Takes personal responsibility for foreseeable consequences</td>
<td>2</td>
<td>7</td>
<td>25</td>
<td>33</td>
</tr>
<tr>
<td>Y10. Takes personal responsibility for unforeseeable consequences</td>
<td>10</td>
<td>20</td>
<td>22</td>
<td>16</td>
</tr>
<tr>
<td>Y11. Willing to act without complete information</td>
<td>20</td>
<td>15</td>
<td>25</td>
<td>7</td>
</tr>
<tr>
<td>Y12. Able to “define the problem” early in the judgment process</td>
<td>2</td>
<td>5</td>
<td>22</td>
<td>38</td>
</tr>
</tbody>
</table>

Possible Qualities Involving the PJPA Situation:

<table>
<thead>
<tr>
<th>Y13. A right or best “answer” exists, whether or not recognized</th>
<th>5</th>
<th>15</th>
<th>32</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y14. Decision is not ideal: it reflects practical compromises</td>
<td>1</td>
<td>11</td>
<td>31</td>
<td>24</td>
</tr>
<tr>
<td>Y15. Decision can be justified (defend, documented)</td>
<td>1</td>
<td>2</td>
<td>12</td>
<td>32</td>
</tr>
<tr>
<td>Y16. Other accountants would agree with the decision</td>
<td>1</td>
<td>19</td>
<td>42</td>
<td>5</td>
</tr>
<tr>
<td>Y17. More than one course of action has something to recommend</td>
<td>6</td>
<td>20</td>
<td>31</td>
<td>10</td>
</tr>
<tr>
<td>Y18. Rules or standards for making the decision do not exist</td>
<td>16</td>
<td>25</td>
<td>23</td>
<td>3</td>
</tr>
</tbody>
</table>

1Categories were assigned the following definitions: (3) the quality is essential to PJPA; (2) the quality is important but not essential; (1) the quality is relevant but not important; (0) the quality is irrelevant to PJPA.

2For each item, the modal category is in italic.

Characterization Results

Table 2 contains the yellow questionnaire results. Most items differed from others in the proportions assigned to various categories, indicating that the pattern of responses is not random. (Of the 153 comparisons possible among the 18 items, 110 were significantly different, p<.05 two-tailed, using simple t-tests on category means.)
Table 3. Characterizations of PJPA (Blue Questionnaire)

<table>
<thead>
<tr>
<th></th>
<th>Number of Respondents Choosing Each Category(^1,2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0  1  2  3</td>
</tr>
<tr>
<td><strong>B1.</strong> &quot;I will try to make complete judgments that will prevent future problems, not just solve the present problem&quot;</td>
<td>2  16 38 13</td>
</tr>
<tr>
<td><strong>B2.</strong> &quot;My judgments are fairly routine, not requiring much real thought&quot;</td>
<td>9  42 16 2</td>
</tr>
<tr>
<td><strong>B3.</strong> &quot;My judgments include an explicit weighting of the consequences of error&quot;</td>
<td>2  21 27 19</td>
</tr>
<tr>
<td><strong>B4.</strong> &quot;When making a judgment, I work out the general principles, then leave the details to other people&quot;</td>
<td>20  34 14 1</td>
</tr>
<tr>
<td><strong>B5.</strong> &quot;Authoritative rules and standards provide useful guidance to me in making judgments&quot;</td>
<td>2  21 35 11</td>
</tr>
<tr>
<td><strong>B6.</strong> &quot;Because I usually know early on what to do, I use information mainly to corroborate my views&quot;</td>
<td>18  30 20 1</td>
</tr>
<tr>
<td><strong>B7.</strong> &quot;Short-run considerations get more weight in my judgments than do long-run considerations&quot;</td>
<td>14  40 14 1</td>
</tr>
<tr>
<td><strong>B8.</strong> &quot;I do not consult anyone else about a judgment until I have thought it entirely through myself&quot;</td>
<td>4  21 28 16</td>
</tr>
<tr>
<td><strong>B9.</strong> &quot;Responsibility for the consequences of my judgments is assigned entirely to me&quot;</td>
<td>9  21 20 19</td>
</tr>
<tr>
<td><strong>B10.</strong> &quot;Whether or not I recognize it at the time, a 'right' or 'best' judgment exists in each case&quot;</td>
<td>3  21 26 19</td>
</tr>
<tr>
<td><strong>B11.</strong> &quot;When making judgments, I am constrained by rules and regulations more than is necessary&quot;</td>
<td>18  47 4  0</td>
</tr>
<tr>
<td><strong>B12.</strong> &quot;In a judgment, I get early advice on the general direction to take, then work out the details&quot;</td>
<td>12  38 15 4</td>
</tr>
<tr>
<td><strong>B13.</strong> &quot;I act on what needs to be dealt with now, choosing future actions as the future develops&quot;</td>
<td>7  38 21 3</td>
</tr>
<tr>
<td><strong>B14.</strong> &quot;During a judgment, my view as to what to do changes as information arrives&quot;</td>
<td>2  29 29 9</td>
</tr>
<tr>
<td><strong>B15.</strong> &quot;I must decide what to do without having all the information I would like to have&quot;</td>
<td>9  39 17 4</td>
</tr>
</tbody>
</table>

\(^1\) Categories were assigned the following definitions: (3) it is true for all, or nearly all, of your judgments; (2) it is true for most of your judgments; (1) it is true for only some of your judgments; (0) it is true for none, or nearly none, of your judgments.

\(^2\) For each item, the modal category is in italic.

The quality viewed as most important to PJPA was that the decision can be justified (item Y15): almost as important was the accountant
or auditor's analytical ability (item Y2). Next in importance were the ability to “define the problem” early in the judgment process (item Y12) and practical experience (item Y5). Least important were the absence of rules or standards for the decision (item Y18), willingness to act without complete information (item Y11), and nonanalytical intuition (item Y6).

There were few differences by country or position, and the main group of interitem correlations involved taking personal responsibility for consequences, prudence, ethics, and theory. We interpreted these to suggest that taking responsibility is ethical and prudent but applies to situations in which theory exists and a right answer can be identified, and that prudence is ethical and implies analysis and knowledge. This interpretation is consistent with the characterization of responsibility for medical judgment described by Bok (1979, Ch. 2).

Table 3 contains the blue questionnaire results. These results were less emphatic than those of the yellow questionnaire. While category 3 (the greatest importance) was the most popular response for seven of the items in Table 2, it was not the modal category in Table 3. Moreover, of the 105 comparisons possible among the 15 items in Table 3, 66 were significantly different ($p<0.05$, $t$-test), compared with 110 of the 153 comparisons reported in Table 2.

The statements that respondents perceived to fit their judgments most of the time were “trying to make complete judgments to prevent future problems” (item B1), “explicitly weighing the consequences of error” (item B3), and “a ‘right’ or ‘best’ judgment exists in each case” (item B10). In addition, item B5, “rules and standards provide useful guidance” and item B8, “consult after I have thought it through,” were also seen to fit well.

Statements fitting respondents' judgments least often were “constrained by rules and regulations more than necessary” (item B11), “work out general principles, leave details to others” (item B4), “give more weight to short-run considerations than long-run” (item B7), “know early what to do, use information mainly to corroborate” (item B6), and “get early advice on direction then work out details” (item B12).

As is the case with the yellow responses, there were only scattered differences by country, position, and years of experience. Also, the main correlation cluster again involved responsibility related to complete solutions, explicit weighing of consequences, and thorough thinking.
Characterization Conclusions

From the two groups of respondents PJPA is characterized as being analytical and fully worked out, and exercised on the basis of practical experience but using rules and standards for guidance. Information is needed to reach the answer and, because judgment must be justified, action should not be taken without full information. Early problem definition, which is desirable, does not imply that the answer is obvious early. The exercise of judgment requires prudence and carries personal responsibility, particularly for consequences that would have been foreseen through analysis and knowledge.

As noted earlier, this characterization is intended to provide a general background about how public accountants themselves see PJPA. It is based on direct but general questions, and as such may reflect what respondents thought PJPA should be more than what it is. There is no information about what kinds of circumstances the respondents had in mind. The next sections of the paper will examine these circumstances in more detail and develop evidence that is not so perceptually direct.

Kinds of Situations Which Involve PJPA

Explanation and Design

In preparation for detailed questions on the judgment process, respondents to the yellow questionnaire were asked to “think back to a decision or action which you recently made on the basis of your professional judgment and which, looking back at it, you feel was made satisfactorily.” Without further prompting, respondents then were asked to describe the situation briefly, specify what event began the process, and what decision or action ended it, and then list (in chronological order) “everything that happened” between those beginning and ending points. (“Everything” was defined as “all the things you did or considered doing, all information you obtained, and all other relevant events.”) The response format provided an identifying number for each event listed by the respondent.

This section deals only with the kinds of situations chosen. (Detailed questions following respondents’ situation choices are referred to in later sections of the paper.) Respondents were asked to choose their own PJPA situations in order to leave them “in control,” thus ensuring that the later questions dealt with PJPA as the respondents saw it. Meeting this objective created two problems.

First, the situations chosen by respondents may not be particularly representative of “ordinary” judgment situations because they might
be affected by factors such as recency, novelty, or a wish to demonstrate good judgment. This potential problem was partly countered by asking (on the next page of the yellow questionnaire) exactly the same questions but for a decision or action made unsatisfactorily. It also was countered, even more strongly, by a section of the blue questionnaire which placed respondents in a particular situation. Further discussion of this issue will be deferred to the section “Further Investigation of Sequence;” but in the meantime the reader should note that the tests of hypotheses in this section and the next are conditional on the situations chosen by the respondents.

Second, respondents might choose situations which “fitted” in some way the questions they were asked. This possibility was forestalled by asking respondents to choose and describe the situations before being asked about them, so that they were committed to their descriptions, and by asking that the descriptions be event-oriented, not feeling- or thought-oriented, in order to increase the objectivity of the recall.

**Situations Hypotheses**

**H1**: PJP will not be limited to auditing situations. This is largely a definitional matter, but the present literature tends to concentrate on audit judgment. It therefore may be instructive to identify the situations in which public accountants themselves think their judgment is required.

**H2**: PJP is largely a process of responding to the demands of the environment, rather than a proactive process in which the accountant initiates the action. This test relates to arguments, developed by Gibbins (1984, for example, pp. 106 and 118–19), indicating that, given the pressures of the environment and the effectiveness of experience-derived response prototypes, PJP usually can be characterized as a responsive process.

**H3**: PJP will be manifested by a series of actions and events, rather than by a single choice. This issue again is definitional but, as argued by Hogarth (1981) and Gibbins (1984), is very important for properly modeling the judgment process and designing research situations (especially in experiments). In a sequence of events, not only is the judgment process itself extended, but there are opportunities to act during the process (including actions to gather more information), making each information-processing step conditional on the results of previous steps and actions. This drawn out, conditional process is not what is modeled by typical experimental or quasi-experimental (“Lens”) research (see reviews by Ashton (1983) and
Libby (1981)), in which participants make simple choices each requiring seconds or minutes, information is presented in one package, and statistical analysis is performed using simultaneous processing models such as multiple regression. Therefore, obtaining support for this hypothesis would suggest the need for different modeling of the judgment process than that done to date.

Situations Results

Of the 67 yellow questionnaire respondents, 66 provided a "satisfactory" PIPA experience, while only 52 provided an "unsatisfactory" one. (The reluctance or inability to provide the latter was not found to be related to position or country.) The 118 situations included investigating accounting disclosure alternatives; examining errors discovered during the audit; giving clients advice on accounting and financial matters; dealing with staff problems; investigating possible fraud; evaluating whether a client is a going concern; and many other circumstances.

Table 4 contains classifications of the "satisfactory" and "unsatisfactory" situations. The data support hypothesis H1 above: more than one half of the situations relate to accounting matters, while only a quarter relate to auditing. The choice of accounting rather than auditing was related to the respondent's position: holders of senior

<table>
<thead>
<tr>
<th>Judgment Situation</th>
<th>Partners</th>
<th>Managers</th>
<th>Supervisors</th>
<th>Assistants</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Satisfactory&quot; situations:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accounting: figures or disclosure</td>
<td>10</td>
<td>9</td>
<td>10</td>
<td>6</td>
<td>35</td>
</tr>
<tr>
<td>Auditing: procedures or opinion</td>
<td>3</td>
<td>3</td>
<td>8</td>
<td>8</td>
<td>22</td>
</tr>
<tr>
<td>Staffing, hours</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Client advice and sundry</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>&quot;Unsatisfactory&quot; situations:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accounting: figures or disclosure</td>
<td>5</td>
<td>9</td>
<td>8</td>
<td>5</td>
<td>27</td>
</tr>
<tr>
<td>Auditing: procedures or opinion</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>Staffing, hours, fee</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Client advice and sundry</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>Total of both kinds of situations:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accounting: figures or disclosure</td>
<td>15</td>
<td>18</td>
<td>18</td>
<td>11</td>
<td>62</td>
</tr>
<tr>
<td>Auditing: procedures or opinion</td>
<td>5</td>
<td>3</td>
<td>10</td>
<td>14</td>
<td>32</td>
</tr>
<tr>
<td>Staffing, hours, fee</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>Client advice and sundry</td>
<td>6</td>
<td>5</td>
<td>0</td>
<td>4</td>
<td>15</td>
</tr>
</tbody>
</table>
positions chose very few auditing situations, while holders of junior positions chose relatively more. (Comparing the accounting and auditing rows of the bottom panel of Table 4, \( \chi^2 = 9.79 \), d.f. = 3, \( p < .025 \).)

Table 5 summarizes the events that respondents perceive to be responsible for starting the judgment process. The table classifies the events into three groups: (1) those initiated by the environment, (2) those initiated by the respondent, and (3) those involving an interaction between the respondent and the environment. (The third group includes, for example, reading through working papers and "realizing there was a problem," or holding a meeting out of which emerges a need to exercise judgment.) The data contained in this table support the second hypothesis (H2) of a largely responsive process: 47 of the 118 events (39.8 percent) were environmental, 58 (57.6 percent) were interactive, and 3 (2.5 percent) were respondent-initiated. The respondents, therefore, saw themselves as dealing with problems as they arose: even the interactive events involved an initial perception of an error or problem. Only three events were clearly pro-active situations in which the respondent chose to begin the process.

<table>
<thead>
<tr>
<th>Events</th>
<th>Satisfactory Result</th>
<th>Unsatisfactory Result</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Environmental events:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phone call to respondent</td>
<td>15</td>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td>Visit to respondent’s office by staff member</td>
<td>6</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>Request by client</td>
<td>6</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>Other environmental events</td>
<td>3</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td><strong>Events arising from an interaction with environment:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discovery of error or problem</td>
<td>9</td>
<td>5</td>
<td>14</td>
</tr>
<tr>
<td>Realization or awareness by respondent of a problem</td>
<td>4</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Audit review or visit to client’s office</td>
<td>16</td>
<td>15</td>
<td>31</td>
</tr>
<tr>
<td>Discussion and meetings</td>
<td>3</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>Assignment of respondent to job or task</td>
<td>2</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td><strong>Events initiated by respondents:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thinking about how to improve job</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Phone call or suggestion by respondent</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td>66</td>
<td>52</td>
<td>118</td>
</tr>
</tbody>
</table>
Table 6 summarizes the number of events respondents listed as part of the judgment process and time taken to formulate the judgments. Hypothesis H3, that PJPA involves a series of actions, is strongly supported. The "satisfactory" situations involved an average of 9.54 events and the "unsatisfactory" situations, 8.23 events. The minimum number of events was 4, and two judgment processes involved 17 events. Many processes took a long time to complete: more than half took days or longer, with 11 taking months to finish.

**Situations Conclusions**

The results provide evidence that PJPA largely is a responsive, sequential process, operating over an extended period of time and involving many events. Issues were found to be broader than audit procedures and techniques, especially for more senior public accountants. These results raise doubt about the applicability of much of the extant research to PJPA as it exists in the field. We admit, however, that if the respondents' chosen situations are biased toward the novel or otherwise nonroutine, our findings imply criticism of the extant research only to the degree that such research bears on nonroutine situations.

<table>
<thead>
<tr>
<th>Table 6. Number of Events and Time Taken in Judgment Process</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of Events and Time Taken</strong></td>
</tr>
<tr>
<td><strong>Satisfactory Result</strong></td>
</tr>
<tr>
<td><strong>Unsatisfactory Result</strong></td>
</tr>
<tr>
<td>Number of events reported by respondent:</td>
</tr>
<tr>
<td>Average number of events</td>
</tr>
<tr>
<td>9.54</td>
</tr>
<tr>
<td>8.23</td>
</tr>
<tr>
<td>Number of times number of events reported:</td>
</tr>
<tr>
<td>4–5 events</td>
</tr>
<tr>
<td>3 times</td>
</tr>
<tr>
<td>9 times</td>
</tr>
<tr>
<td>6–10 events</td>
</tr>
<tr>
<td>42 times</td>
</tr>
<tr>
<td>31 times</td>
</tr>
<tr>
<td>11–15 events</td>
</tr>
<tr>
<td>20 times</td>
</tr>
<tr>
<td>12 times</td>
</tr>
<tr>
<td>16–17 events</td>
</tr>
<tr>
<td>2 times</td>
</tr>
<tr>
<td>0 times</td>
</tr>
<tr>
<td>67</td>
</tr>
<tr>
<td>52</td>
</tr>
<tr>
<td>Reported time taken by judgment process:</td>
</tr>
<tr>
<td>Seconds</td>
</tr>
<tr>
<td>0 times</td>
</tr>
<tr>
<td>0 times</td>
</tr>
<tr>
<td>Minutes</td>
</tr>
<tr>
<td>7 times</td>
</tr>
<tr>
<td>5 times</td>
</tr>
<tr>
<td>Hours</td>
</tr>
<tr>
<td>15 times</td>
</tr>
<tr>
<td>7 times</td>
</tr>
<tr>
<td>Days</td>
</tr>
<tr>
<td>31 times</td>
</tr>
<tr>
<td>25 times</td>
</tr>
<tr>
<td>Weeks</td>
</tr>
<tr>
<td>7 times</td>
</tr>
<tr>
<td>7 times</td>
</tr>
<tr>
<td>Months</td>
</tr>
<tr>
<td>6 times</td>
</tr>
<tr>
<td>5 times</td>
</tr>
</tbody>
</table>
Sequence of Events in PJPA

Explanation and Design

After a respondent had written down the events that happened during the judgment situation he/she had chosen, questions were asked about the way those events unfolded and about the environmental influences on them. This section focuses on a series of questions relating to the ordering and importance of certain events in the sequence. The questions centered on three points in the event sequence: (1) the point at which the nature of the judgment problem was clearly defined, (2) the earliest point at which the respondent felt a preference about what the ‘final decision or action should be,’ and (3) the earliest point at which the respondent felt the ‘final decision or action could be adequately justified or defended.’

Sequence Hypotheses

This section tests several hypotheses developed from the PJPA characterization proposed by Gibbins (1984), which is similar to that outlined in Waller and Felix (1984a). In this characterization, public accountants' experience leads them to develop response prototypes which are available in memory and which can be applied very rapidly to the situation as perceived. The response prototype is triggered by cues (not necessarily very many of them) in the situation and, because the prototype is available, it comes to mind very quickly. This rapidly rising response preference is evaluated against perceived risks and constraints before being implemented. A major component of such an evaluation is to ensure that any action can be justified.

H4: A response preference is felt early in the judgment process. Operationally, we hypothesize that, about as soon as the public accountant has perceived enough about the situation to define what the judgment problem is, he/she will have a response preference. (This operationalization also follows from our previous finding (that is, that defining the problem early in the process is important).)

H5: The public accountant will feel a response preference before he/she is able to justify that response. This hypothesis and H4 imply that much of the information gathering during the judgment process is in the service of providing the justification rather than of determining a preferred decision or action.

H6: The point at which the final decision or action can be justified will be perceived as the most important event in the judgment process. This hypothesis is derived from the earlier result that respondents rated justification as the quality most important to PJPA.
H7: A more experienced public accountant develops response preferences faster than a less experienced accountant. This hypothesis is based on the argument that experience has value, partly because it produces response prototypes which are more developed and more quickly accessed (see Gibbins 1984, p. 109).

H8: During the judgment process, actions which reduce future ability to respond will be avoided, when possible. If judgment is a sequential process, proceeding by incremental steps, each conditional on information arriving since the preceding step, it will be to the accountant's advantage to keep his/her options open, and avoid being committed to later actions until the process is completed.

Sequence Results

Most of the results related to the sequence hypotheses are set out in Table 7. The upper portion of the table shows that, on average, the judgment problem was defined earlier than the point in time when the respondent felt a preference for the final decision/action. Further, decision/action preference was felt earlier than the point at which the decision/action could be adequately justified or defended. Specifically in situations that turned out satisfactorily, respondents felt that the problem was defined by 32 percent of the way through the process, felt a preference about the final decision 37 percent of the way through, and felt that the action could be justified 48 percent of the way through. In situations which turned out unsatisfactorily, the three points occurred 45 percent, 50 percent, and 60 percent of the way through the process.

Data contained in Table 7 support H4. For both kinds of situations, the preference is felt early, well before half way through the process. Any delay is associated with the time it takes to define the problem: a preference is felt very soon after that. The earliest preference and problem-defined points were almost interchangeable, as Panel (A) in the middle of Table 7 shows. Definition and preference coincided 21 times (12 + 9), definition preceded preference 51 times (30 + 21), and preference preceded definition 44 times (24 + 20). However, the frequencies of 51 and 44 are not significantly different according to the sign test.

Evidence of early preference for the decision/action was also obtained from a later question, which asked, "Throughout the judgment, was it clear to you what the final decision or action should be?" The 118 responses were: strongly no, 2; no, 56; yes, 52; strongly yes, 8. Thus, about half the respondents felt a preference all along.
### Table 7. Judgment Sequence Questions

<table>
<thead>
<tr>
<th>Event</th>
<th>Satisfactory Result</th>
<th>Unsatisfactory Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average number of events in judgment process</td>
<td>9.54</td>
<td>8.23</td>
</tr>
<tr>
<td>Average perceived point at which three major events occurred (bracketed figures are the averages of the proportion of each respondent's total events at which he/she placed each event):</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Nature of judgment problem clearly defined</td>
<td>4.06 (.32)</td>
<td>3.47 (.45)</td>
</tr>
<tr>
<td>2. Earliest point at which a preference about final decision/action was felt</td>
<td>4.64 (.37)</td>
<td>4.16 (.50)</td>
</tr>
<tr>
<td>3. Earliest point at which final decision/action could be adequately justified</td>
<td>5.97 (.48)</td>
<td>5.11 (.60)</td>
</tr>
</tbody>
</table>

**Comparison of the three events' sequencing:**

(A) Event 1 vs. event 2:
- Defined (1) before preference (2): 30 times 21 times
- Defined in same event as preference: 12 times 9 times
- Preference (2) before defined (1): 24 times 20 times
- Incomplete answers: 1 time 2 times

(B) Event 2 vs. event 3:
- Preference (2) before justifiable (3): 38 times 27 times
- Preference in same event as justifiable: 24 times 17 times
- Justifiable (3) before preference (2): 5 times 5 times
- Incomplete answers: 0 times 3 times

(C) Event identified as crucial to the way the judgment process turned out:
- Event earlier than any of these three: 5 times 4 times
- Definition (1) identified as crucial: 12 times** 7 times**
- Preference (2) identified as crucial: 17 times** 9 times**
- Justification (3) identified as crucial: 20 times** 9 times**
- Event later than any of these three: 10 times 12 times
- Other event patterns identified: 12 times 10 times
- No event identified as crucial: 7 times 6 times

** Some double counting when definition and/or preference and/or justification events coincided.

(This preference was stronger in satisfactory than unsatisfactory situations: \( \chi^2 = 7.6, \ p < .01 \))

H5 is more strongly supported. The response preference was usually felt before the decision/action could be justified, and hardly ever after: Panel (B) of Table 7 shows that the two points coincided 41 times (24 + 17), preference preceded justification 65 times (38 + 27)
and justification preceded preference 10 times (5 + 5). Additionally, these frequencies are significantly different according to the sign test ($p = .000$).

Panel (C) of Table 7 does not provide much support for H6. When respondents were asked to indicate which, if any, events were “particularly crucial to the way the judgment process turned out,” as hypothesized, the justification point was mentioned most frequently. But other points also were cited, so that “justification” does not have a sharp lead.7

A variety of definitions of what is crucial is also implied by the results of a later question. Recall that in the earlier “Characterization” section of the paper, taking responsibility for consequences was prominent. A later question asked, “Was responsibility for any consequences clearly assigned to you?” The 119 responses were strongly no, 8; no, 45; yes, 54; strongly yes, 12. (These proportions were not statistically different for satisfactory and unsatisfactory situations.) Thus, perceptions of individuals’ responsibility for the recalled judgment situation were diverse.

H7 predicts a positive correlation between the respondent’s experience and rapidity of developing response preferences. Though this prediction involves the speed of cognitive processes rather than, necessarily, events, it seems reasonable to expect experience to correlate (negatively) with the proportion of events that have happened by the time the respondent feels a response preference. No significant correlation appeared in these data, however, for either satisfactory or unsatisfactory situations. The hypothesis is, therefore, not supported.

A later question was asked to determine the respondent’s experience with the specific judgment situation he/she chose: “Was the judgment one you were accustomed to making?” The 118 respondents were: strongly no, 4; no, 46; yes, 62; strongly yes, 6. The proportion of “yes” responses is higher in the satisfactory than in the unsatisfactory situation: $\chi^2 = 6.84$, $p < 0.01$. However, this specific experience also did not correlate with the point in the sequence of events at which a response preference was felt. Thus, no support for H7 appears here either.

Data relating to H8 were obtained by asking respondents to identify whether “any of the things you did or learned prior to the final decision or action committed you to a later course of action or otherwise limited your options.” Of the 119 satisfactory and unsatisfactory situations, 72 respondents indicated no such events, 17 indicated 1 event, 17 indicated 2 events, 7 indicated 3 events, and 6
indicated 4 or more events. The hypothesis is supported: little, if any, commitment prior to the final decision/action was perceived.

*Sequence Conclusions*

The results generally support the model of preferences developing rapidly once the situation is understood, followed by a determination that the preferred decision/action can be justified. Experience effects did not appear in the speed of developing preferences, nor was justification clearly established as the most crucial stage in the judgment process.

*Further Investigation of Sequence*

*Explanation and Design*

In the preceding sequence, the yellow questionnaire respondents were asked to choose circumstances that had required the exercise of professional judgment. Although this approach had the advantage of permitting natural variation in the judgment circumstances, it also may have limited the results in some ways.

One obvious limitation is that the respondents were asked to recall professional judgment circumstances. Quite aside from the serious question of whether people have conscious access to their cognitive processes (Gibbins 1984; Nisbett and Wilson 1977; Nisbett and Ross 1980), such recall will include many factors in addition to cognition. The Gibbins (1984) model proposes that response preferences develop early in the judgment process because they are based on learned response prototypes available in memory when the situation arises. But, the role of these prototypes may be obscured by all the other specifics of the recalled situation. Asking respondents to anticipate a situation should show that role more clearly, however, because thinking ahead is likely to be based on any response preferences that are available in memory.

Another potential limitation (raised earlier), which might result from asking respondents to suggest the circumstances, is bias toward producing more “interesting” (for example, novel) situations. In particular, respondents are likely to select unusual cases to discuss, and thus, neglect the more routine or structured problems. According to Gibbins (1984), any response prototypes would be better developed for standard than for nonstandard situations. If respondents were asked to deal with a recognized problem, the results should reflect responses they already were prepared to make.

In addition, asking each respondent to choose a situation with
which he/she was recently familiar may have obscured the effects of position or experience. The respondent's familiarity with the chosen situation, and the strength of any response prototype available for that situation, need not be related to position or general experience. A partner and a junior might choose different situations, but be similarly experienced in the particular situations they chose (for example, each having encountered his/her situation several times recently).

In recognition of these limitations, and to take advantage of the split questionnaire design by providing a partial replication of the foregoing sequence study, respondents to the blue questionnaire were asked to think prospectively about a standard situation common to all. The blue questionnaire respondents were asked to think of a familiar client, and were then given the following judgment process starting point: "Your audit team discovers that an employee of the client who was fired (discharged) six months ago is still on the payroll." The respondents then were asked to describe how they would proceed with the judgment, up to and including a plausible final decision or action. After providing a written answer these respondents were asked the questions asked of the respondents to the yellow questionnaire. As these questions referred to an expected pattern (rather than the past pattern of the yellow questionnaire), expectations wording was used where necessary; otherwise an identical wording to the yellow questionnaire was used.

A modification of the events listing used in the yellow questionnaire was needed. Respondents could not be asked to list the events that had happened because they were projecting what they would do upon encountering the problem. It was argued earlier that the judgment process is conditional upon any information which arrives during the process. If this were valid, then asking a respondent to list his/her actions would require him/her to make assumptions about what would be learned as the actions are carried out. To clarify, these respondents were asked to list both "what you would do" and "your assumptions about plausible results of what you do, or other plausible events that you would expect, to reflect the information you would require for your judgment to proceed." To explain the answer format (a list of items in sequence, with each being identified as either what the respondent would do or what he/she assumed), a humorous example was given of how a person might respond to a spouse's asking if he/she would like to go to a movie.⁹
Further Sequence Investigation Hypotheses

Four additional hypotheses (specified below) will be tested as replications and extensions of the earlier results.

H9: A response preference will be expected early in the sequence of events (replication of H4, which was supported earlier).

H10: The response preference will be expected to be felt before it is clear that a response can be justified (replication of H5, which was supported earlier).

H11: The justification point will be expected to be the most important event in the judgment process (replication of H6, which was not supported earlier).

H12: Actions which reduce future ability to respond will be avoided (replication of H8, which was supported earlier).

No hypothesis is offered with respect to experience or position effects; instead, any such effects are reported as descriptive evidence. While the familiarity with the given auditing payroll situation may differ by position, it was chosen as one that could plausibly be answered by most public accountants. (Note that the experience hypothesis (H7) was not supported in the earlier analysis.)

Further Sequence Investigation Results

Sixty-two respondents wrote down what they would do and answered the subsequent questions. A summary of their responses is set out in Table 8, while the Chart presents a comparison of the Table 7 (yellow questionnaire) results with those of Table 8 (blue questionnaire). The respondents expected that an average of 7.58 actions would be made to deal with the problem. Those actions plus the event-starting process give a mean of 8.58 events, which is very similar to the average of 8.97 events reported retrospectively by the yellow questionnaire respondents (8.97 is the weighted average of 9.54 and 8.23 in Table 7). An average of 4.45 assumptions were expected to be made about results of actions, giving a total expected average of 13.03 steps to deal with the payroll problem. There were no differences by position in the expected number of actions, number of assumptions, or total number of steps.

The results also support H9. On average, a response preference is expected to be felt by the 5.15 event (41 percent of the way through the process), even before the problem is defined clearly, which is expected to occur at the 6.94 event (54 percent of the way through). Referring to Panel (A) of Table 8, the preference was felt before the problem was defined 32 times, versus only 16 times in the opposite
Table 8. Results from Further Investigation of Judgment Sequence

<table>
<thead>
<tr>
<th>Event</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average number of events expected in process</td>
<td></td>
</tr>
<tr>
<td>Event given to respondents to start process</td>
<td>1.00</td>
</tr>
<tr>
<td>Actions respondents said they would take</td>
<td>7.58</td>
</tr>
<tr>
<td>Events involving respondents' assumptions about results of their actions</td>
<td>4.45</td>
</tr>
<tr>
<td>Average perceived point at which three major events expected (bracketed figures are the averages of the proportion of each respondent's total events at which he/she placed each event):</td>
<td></td>
</tr>
<tr>
<td>1. Nature of judgment problem clearly defined</td>
<td>6.94 (.54)</td>
</tr>
<tr>
<td>2. Earliest point at which a preference about final decision/ action was expected</td>
<td>5.15 (.41)</td>
</tr>
<tr>
<td>3. Earliest point at which final decision/action expected to be adequately justified</td>
<td>8.19 (.61)</td>
</tr>
</tbody>
</table>

Comparison of the three events' sequencing:

(A) Event 1 vs. event 2:
- Defined (1) before preference (2)                                  16 times
- Defined in same event as preference                                 13 times
- Preference (2) before defined (1)                                  32 times
- Incomplete answer                                                  1 time

(B) Event 2 vs. event 3:
- Preference (2) before justifiable (3)                              42 times
- Preference in same event as justifiable                            13 times
- Justifiable (3) before preference (2)                              7 times

(C) Event identified as crucial to the way the judgment process turned out:
- Event earlier than any of these three                              7 times
- Definition (1) identified as crucial                               20 times**
- Preference (2) identified as crucial                               11 times**
- Justification (3) identified as crucial                            24 times**
- Event later than any of these three                                3 times
- Other event patterns identified                                     11 times
- No event identified as crucial                                     4 times

** Some double counting when definition and/or preference and/or justification events coincided.

direction ($p = .015$ by the sign test). Thus, the two points were interchangeable in the retrospective situation, but not so in the prospective one.

As the Chart shows, the earliest preference, at 41 percent of the way through, is about the same as the 43 percent (average of satisfactory and unsatisfactory situations) for the retrospective situation. Conversely, respondents expected to take longer to define the problem (54 percent in the prospective, versus the 38 percent average
Chart. Comparison of Judgment Sequence Patterns

Notation used below: B — beginning of judgment process
D — point at which problem is clearly defined
P — earliest point action preference felt
J — earliest point justification is adequate
A — decision/action

Yellow questionnaires (satisfactory and unsatisfactory combined)

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>D</th>
<th>P</th>
<th>J</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion of total number of events at which point occurs</td>
<td>0</td>
<td>.38</td>
<td>.43</td>
<td>.53</td>
<td>1.0</td>
</tr>
<tr>
<td>Percentage of times point is indicated as crucial</td>
<td>before</td>
<td>7</td>
<td>15</td>
<td>20</td>
<td>23</td>
</tr>
</tbody>
</table>

Blue questionnaires

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>P</th>
<th>D</th>
<th>J</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion of total number of events at which point occurs</td>
<td>0</td>
<td>.41</td>
<td>.54</td>
<td>.61</td>
<td>1.0</td>
</tr>
<tr>
<td>Percentage of times point is indicated as crucial</td>
<td>before</td>
<td>9</td>
<td>26</td>
<td>14</td>
<td>32</td>
</tr>
</tbody>
</table>

(Data for this chart are from Tables 7 and 8)

of the satisfactory and unsatisfactory situations in the yellow questionnaire).

H110 also is supported: a preference is expected much sooner than justification is expected (8.19 event, 61 percent of the way through). By Panel (B) of Table 8, preference was expected to precede justification 42 times, versus only 7 times in the opposite direction \( p = .000 \) by the sign test. This is a stronger effect than that which was obtained for the yellow questionnaire.

Panel (C) of Table 8 provides the same degree of weak support for H11 as was earlier provided for H6. Though justification is the most commonly cited crucial event, it does not dominate.

Respondents also were asked to indicate whether anything they expected to do or learn would commit them to a later course of action or otherwise limit their options. Their results do not, however, provide much support for H12 (that is, in contrast to the strong support reported earlier for H8 in the yellow questionnaire). Of the
62 respondents, 20 indicated no such events, 10 indicated 1 event, 11 indicated 2 events, 10 indicated 3 events, and 11 indicated 4 or more events. Less than a third indicated no constraints, compared with more than 60 percent (72 of 119) in the yellow questionnaire. Respondents did not feel constrained, therefore, when looking back, but when looking forward they expected to be constrained.10

Further Sequence Investigation Conclusions

The general model of rapidly developing preferences was supported again, using a different group of respondents and a prospective rather than a retrospective situation. Additionally, justification again was cited as crucial most often, but did not dominate. Finally, the results continued to show no effects of years of experience or position in the accounting firm.

Consultative Environment of PJPA

Explanation and Design

In the previous sections, we focused on the individual accountant’s relationship to the judgment situation. Though there have been references to the external environment via factors such as personal responsibility for consequences, explicit attention should be paid to the social/organizational context within which PJPA operates. PJPA is of course carried out in public accounting firms, so PJPA research should recognize its organizational setting (Gibbins 1983; Gibbins and Wolf 1982a and 1982b; Mock and Turner 1981). Research on auditors’ judgment has begun to examine the effects of group and organizational factors. For example, Pratt and Jiambalvo (1981) considered leadership variables within the audit team, and Abdelkhalik, Snowball, and Wragge (1983) showed that external auditors’ judgments of internal auditing systems were affected by the internal auditors’ level of reporting responsibility.

The design of the yellow questionnaire considered some aspects of the organizational context, one of which is the consultative environment. During the questionnaire-development interviews, frequent references were made to working paper reviews, concurring reviews, requirements for checking with advisors or industry specialists, and so on. In a study of auditor probability assessments, Solomon (1982) provided evidence that consultation improves individual assessments; if audit team members interacted, their assessments tended to be better than aggregates of individual assessments. Two kinds of questions investigating the consultative environment will be reported here: a group of questions related to the specific judgment situation chosen
by the respondent and a general question about the consultation pattern.

On the yellow questionnaire pages, following their listings of events in the chosen "satisfactory" and "unsatisfactory" judgment situations, the respondents were asked to "give an indication of the environment you experienced" in the two situations. One question asked generally whether the final decision or action was affected by consultations with other people. The response scale had four points: strongly no, no, yes, and strongly yes. Four questions asked about the impact of other people in the public accounting firm: the respondent's superiors, colleagues/peers and superiors, and technical advisers. Each of these had a seven-point response scale, 0 indicating the people concerned had had no impact, +1, +2, or +3 indicating degrees of assistance with the judgment, and −1, −2 or −3 indicating degrees of impairment of the judgment.

Two general hypotheses are being investigated here. The first, (H13), following from Solomon (1982) and the foregoing comments, states that consultation will have an important influence on PJPA. The second, (H14), states that the influence will exist at several stages in the process, but will be strongest after the individual accountant has developed a preference about what he/she would like to do. This second hypothesis follows from the psychological model in Gibbins (1984), in which it is proposed that preferences develop quickly, in most cases, and are appraised (for example, via consultation) prior to action.

Consultative Environment Results

Table 9 contains the results of the five questions. On the general question, respondents indicated by a large majority that consultation had affected the decision/action choice: for the satisfactory situations, 54 said "yes" or "strongly yes," against 13 "no" or "strongly no"; for the unsatisfactory situations, 42 were on the "yes" side, versus 11 on the "no" side. In both cases the difference between the "yes" and "no" proportions is highly significant by the sign test ($p = .000$). The proportions did not differ significantly between the satisfactory and unsatisfactory situations, but when asked about specific other people, the results show a significant difference between the two kinds of situations (i.e., persons at all four functional levels were perceived as more helpful in satisfactory than in unsatisfactory situations).

The correlations in Table 9 show some interesting patterns. For the satisfactory situations, the more senior or experienced the re-
Table 9. Consultative Environment of Respondents' Chosen Situations

<table>
<thead>
<tr>
<th>Number of Respondents Giving Answer Indicated</th>
<th>Satisfactory Result</th>
<th>Unsatisfactory Result</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Strong No Yes</td>
<td>Strong No Yes</td>
</tr>
<tr>
<td>Decision/action affected by consultations</td>
<td>1 12 32 22</td>
<td>0 11 25 17</td>
</tr>
<tr>
<td>Subordinates</td>
<td>3 42 22 7 40 6</td>
<td></td>
</tr>
<tr>
<td>Colleagues/peers</td>
<td>3 35 29 6 33 14</td>
<td></td>
</tr>
<tr>
<td>Superiors</td>
<td>2 25 40 18 21 14</td>
<td></td>
</tr>
<tr>
<td>Technical advisors</td>
<td>1 40 26 4 39 10</td>
<td></td>
</tr>
</tbody>
</table>

[Using the full range of the 7-point scale, −3 to +3, two-tailed t tests were done on the response differences in the four questions as between satisfactory and unsatisfactory situations. The t for subordinates is 3.22 (p<.005), for colleagues/peers is 2.57 (p<.02), for superiors is 4.82 (p<.001) and for advisors is 2.67 (p<.02).]

Pearson Product-Moment Correlations

<table>
<thead>
<tr>
<th>Respondent's:</th>
<th>Satisfactory Result</th>
<th>Unsatisfactory Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rank/position*</td>
<td>.13 .26 -.36 .26 -.19 -0.05 .01 -.03</td>
<td></td>
</tr>
<tr>
<td>Experience**</td>
<td>.14 .31 -.21 .27 .02 .00 -.01 .09</td>
<td></td>
</tr>
</tbody>
</table>

Other people:

| Subordinates | .13 .07 -.04 -.07 .13 .17 |
| Colleagues/peers | .10 .10 | .56 .49 |
| Superiors     | .17     |

[Correlations over .25] are significant at p < .05, two-tailed; over .30 at p < .02; over .33 at p < .01; over .35 at p < .005; and over .41 at p < .001.]

* Rank coded as: partner 4, manager 3, supervisor 2, junior 1.
** Experience is in number of years in public accounting.

Respondent, the more helpful colleagues and technical advisers were perceived to be, while superiors were perceived to be less helpful (the 40 instances of superiors rated helpful came primarily from the
less senior people). This finding probably reflects partners’ status as members of the firms and their earlier-stated belief about taking responsibility for judgment results. Subordinates were perceived as being less helpful than other people, and there was no disagreement about that aspect as a function of rank or experience.

These correlations are substantially different, however, for the unsatisfactory situations. Not only are other people perceived to be less helpful, but those perceptions are unrelated to rank or experience. Instead, a strong set of correlations among the helpfulness of colleagues, superiors, and advisers appears for the unsatisfactory situations although it was absent for the satisfactory ones. The perception seems to be that in satisfactory situations, the other people (except subordinates) are not only helpful but each adds independently to the judgment, while in unsatisfactory situations, they are less helpful and when they do affect the judgment, they tend to reinforce each other (presumably to the perceived detriment of the respondent).

The results support H13, indicating that consultation is perceived as important in general, and that specific “others” are helpful. However, each group except superiors is perceived to be “not relevant” more often than “helpful,” indicating that consultation, like other aspects of PJPA, is a complex matter.

The pattern of consultation within the PJPA process was examined by asking respondents to indicate how frequently they discussed or consulted at three times: before having a preference about the decision/action, after having that preference but before acting, and after acting. Respondents were asked to allocate 100 percent of occasions experienced among eight categories: no consultation at any time, consultation only at any one of the three times, consultation at any of the three pairwise combinations of the three times, and consultation at all three times.

Table 10 contains the results of the consultation pattern question, which supports the general importance of consultation: 93.29 percent of the occasions involved consultation and only 6.72 percent did not. Consultation happened during one of the three time periods 45.65 percent of the time, during two periods 34.82 percent of the time, and during all three periods 12.82 percent of the time. To these respondents, PJPA is clearly a consultative, not an individual, process.

Support is shown for H14; the hypothesis that consultation is most important between the time a preference is felt and the time an action is taken (preference screening). When consultation happens only in one time period, it is in that period more than in the other two periods combined (27.51 percent versus 11.45 percent + 6.69
Table 10. General Pattern of Consultation During Judgment Process

| A. | Consultation only during time prior to knowing the preferred decision/action | 11.45 |
| B. | Consultation only during time after knowing preferred decision/action but before acting | 27.51 |
| C. | Consultation only after acting | 6.69 |
| Percentage involving consultation during one time period | 45.65 |
| D. | Consultation during A and B but not C | 21.02 |
| E. | Consultation during A and C but not B | 4.08 |
| F. | Consultation during B and C but not A | 9.72 |
| Percentage involving consultation during two time periods | 34.82 |
| G. | Consultation during all three times | 12.82 |
| Percentage involving consultation at some time | 93.29 |
| H. | No consultation during any of the three times | 6.72 |
| | | 100.01 |

Calculation of Percent of Occasions Involving Consultation That Each of the Three Times Is Presented

<table>
<thead>
<tr>
<th>Prior to knowing preference</th>
<th>Aggregate Involvement</th>
<th>Percent of Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>That time period only (A above)</td>
<td>11.45</td>
<td>7.4</td>
</tr>
<tr>
<td>That time as well as others (D + E + G above)</td>
<td>37.92</td>
<td>24.7</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>49.37</strong></td>
<td><strong>32.1</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Between preference and action</th>
<th>Aggregate Involvement</th>
<th>Percent of Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>That time period only (B above)</td>
<td>27.51</td>
<td>17.9</td>
</tr>
<tr>
<td>That time as well as others (D + F + G above)</td>
<td>43.56</td>
<td>28.3</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>71.07</strong></td>
<td><strong>46.2</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>After action</th>
<th>Aggregate Involvement</th>
<th>Percent of Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>That time period only (C above)</td>
<td>6.69</td>
<td>4.4</td>
</tr>
<tr>
<td>That time as well as others (E + F + G above)</td>
<td>26.62</td>
<td>17.3</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>33.31</strong></td>
<td><strong>21.7</strong></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>153.75</strong></td>
<td><strong>100.0</strong></td>
</tr>
</tbody>
</table>

percent). Overall, the lower panel of Table 10 shows that this time period is involved in 46.2 percent of the consultations, versus 32.1 percent prior to establishing the preference, and 21.7 percent after having acted. Only 7.4 percent of the time does consultation happen
before the preference is known. It appears then that consultation generally happens with the awareness of what the public accountant already would like to do. The potential for inductive use of information in consultation (testing of a hypothesized solution (Cohen 1977; and Chesley 1984)) is obvious, especially when, as demonstrated earlier, preferences seem to arise quickly in individuals’ judgment processes.

Consultative Environment Conclusions

The respondents indicated that consultation is very important to PJPA, particularly consultation with peers, superiors and technical advisors, and particularly after knowing what decision/action is preferred but before acting. These results suggest that research that forces participants to use their own judgment without consultation significantly interferes with the judgment process and is lacking a significant component of external validity.

GENERAL CONCLUSIONS

The objective of this paper has been to improve the specification of the nature of professional judgment in public accounting and to test some hypotheses about how PJPA operates. The questionnaire used was designed to leave the respondents in control of the portrayal so that the results would reflect the natural process as they experience it. The questionnaire was based on a cognitive learning model (Gibbins 1984) and was sent to one person at each of four “levels” in several public accounting firms in five countries. The recipients were a very conscientious group: not only was the overall response rate high, but all sections of the questionnaire were answered by virtually all respondents. This indicates a serious approach to the material by all the respondents and supports the fundamental meaningfulness of the findings. It is admitted, however, that because of the lack of control over the circumstances in which respondents answer and the problems of reactivity (demand characteristics) discussed earlier, a questionnaire instrument is only a suggestive device for developing descriptions and testing hypotheses. Moreover, although several public accounting firms participated in the study, they were not randomly selected and their employee-respondents were obtained on a volunteer basis. Accordingly, the results presented herein, though based on a broad international sample and representing various levels of experience, may not be representative of public accountants in general.

The results indicate that PJPA is a complex process, going well beyond auditing and interacting with the public accountant’s envi-
environment in many ways. Problem definition, justification, keeping one's options open, taking responsibility, and consultation with others are important components of the process. These are fairly general components, with only consultation and taking responsibility showing any relationship to the respondent's position or length of experience.

Several hypotheses based on the propositions in Gibbins (1984) were supported. PPA appears to be largely responsive (rather than pro-active), it involves a series of events (each conditional on preceding events), response preferences arise early in the process (and may therefore condition subsequent information use), and consultation and checking the justifiability of the preference generally precede action. There also is much variation around this general characterization: the process is strongly rooted in the particular situation and environment experienced by the respondent.

We have argued that PPA research should be firmly grounded in the natural environment in which judgment processes operate and develop. The accounting/auditing judgment research literature has shortcomings in this respect: auditing situations are over-emphasized, simple choices are presented (rather than as part of sequences of choices or events), the objective of information gathering is not explicitly considered (for example, problem definition versus preference determination versus justification screening), and the participants are usually uprooted from their natural environment, and its motivations, and are not able to carry out such crucial steps as consulting with others.

NOTES

* This work has been supported financially by Deloitte Haskins and Sells Canada, the Social Sciences and Humanities Research Council of Canada and the University of British Columbia. We are especially grateful to the public accountants from five countries who provided the data reported in this paper.


2. A few descriptive studies have been done, for example, Biggs and Wild (1983); Gibbins and Wolf (1982a); and Hylas and Ashton (1982). However, these are oriented to external auditing in general, not to judgment in particular. Also, professional judgment in public accounting may include issues-auditing, such as accounting policy, disclosure, personnel management, business risk, and client development.

3. Some sections of the questionnaire related to topics beyond this paper's focus and so will not be reported here.
4. Procedures to estimate nonresponse bias were not undertaken because response rates were uniformly high, except for the less enthusiastic Australian response. (Contact with the participating firms was not as direct in Australia as in the other countries and probably led to the lower response rate there.)

5. This was the only place in the questionnaire where the word “judgment” sometimes implied an action rather than a cognitive process.

6. As noted earlier, assumptions about population characteristics are avoided in the statistical tests by using simple chi-square and sign tests.

7. Perhaps the question was not worded well. Panel (C) suggests that a variety of definitions of “crucial” may have been invoked, including assistance in defining the problem and developing the response preference.

8. This starting point was chosen after consultation with public accountants (excluding the respondents) about what sort of situation could plausibly involve public accountants from any level in the firms.

9. An example totally unrelated to the given payroll problem starting point was chosen in order to avoid biasing respondents’ answers.

10. There were no significant differences in the foregoing patterns by position or years of experience.

REFERENCES


———, “After the Audit: Great Expectations or Post-Audit Blues?” Camagazine (June 1982b), pp. 46–51.

R. M. Hogarth, “Beyond Discrete Biases: Functional and Dysfunctional


Evidence on the Nature of Professional Judgment in Public Accounting: Discussion

Gibbins and Emby consider an important problem in the analysis of professional judgment. How do the environment and decision requirements of public accounting interact with the cognitive processes of the accounting expert to produce what we call professional judgment? Studies of human judgment, in general, and expert judgment, in particular, fill many pages of journals in the various social sciences. Much of the judgment research in auditing has used this general research as a base and has attempted to apply the general models and results in the audit context. Findings of such studies, generally, show that many of the results which obtain in other settings do not do so in the auditing context. Clearly the optimality of any decision approach, model, or heuristic depends on the environmental conditions under which the decision is made. The Gibbins and Emby paper is an attempt to identify the environmental conditions peculiar to public accounting and, thereby, further understanding of the judgment processes of the professional accountant. To the extent that the paper attempts to shed light on this important issue, it is a well-motivated piece of research.
Aside from having an important question to address, a good piece of empirical research must develop (or choose) and justify a strong research design, which includes collection of reliable data pertinent to the issue of interest, careful analysis of these data, and powerful statistical tests of the hypotheses developed. It is these questions of experimental design and statistical analysis to which I shall address my comments. Although the paper considers an important problem, there are several weaknesses in the design and statistical analysis which make the reported results less reliable and valid than they otherwise might be. Issues related to questionnaire design and the experimental hypotheses will be considered first. Subsequently, I discuss the statistics used to describe the data and test the hypotheses.

The use of a self-administered questionnaire involves many potential problems (of which the authors are aware), such as attempts to provide answers which meet the questioner's expectations, attempts to keep responses consistent, framing biases, order effects, and so on. Gibbins and Emby raise several of these issues as potential problems, but they do not explain what if anything was done to counteract these effects nor do they analyze data to determine whether or not these issues were a problem in their study. It appears that some additional modifications in the questionnaire design could have moderated these biases or allowed an ex post determination of whether they affected the results. For example, in the yellow version of the questionnaire almost all of the traits or qualities presented could be considered positive character traits, whether or not they apply to professional judgment in public accounting (PJPA). This list might give one the idea that someone thinks these qualities are important to PJPA even if an individual respondent did not think so. Consequently, the respondent might bias his/her response upward. As a check against such a bias, the list could easily have included some qualities that were not considered important by professionals. Similarly, order effects could have been controlled by ordering the items in the questionnaire differently for different subgroups. Framing effects could have been mitigated by rewording some of the statements from the positive to the negative form in some of the questionnaires. For example, one statement included in the yellow questionnaire is "A right or best answer always exists ...". This statement might also read "There is not always a right or best answer ...". These suggestions represent minor modifications in the questionnaire design which would have allowed the authors to enhance the validity of their results.

The authors use a four-point, discrete rating scale to elicit the
subject-auditors importance perceptions of various qualities or statements to professional judgment. The choice of a four-point rating scale, however, is not discussed. Additionally, when this type of psychological-scaling approach has been employed in prior research, typically either a seven- or nine-point scale has been adopted. There is some research in this area that suggests that the inclusion of a middle-scale value may have an important effect on the results. The reasons that a four-point scale was chosen at least should have been discussed by the authors. The use of a discrete scale rather than a continuous scale also is a factor which should have been discussed. A continuous scale similar to the one used by Ashton in his dissertation would make the statistical analysis of the data much simpler; if there were important reasons why a discrete scale was preferred in this case, they should have been presented in the paper.

The judgment description portion of the questionnaire is open-ended. The authors are aware of the problems created in asking subjects for retrospective self reports, and I shall not reiterate them here. However, I think that a discussion of the implications of such factors as rationalization or selective memory for the reported results should have been provided. If the subjects' reports are influenced by these factors, are the results still valid?

I would now like to consider the hypotheses the authors test. Some of these hypotheses are loosely derived from the cognitive learning model propositions presented in Gibbins (1984), while others do not have a clear source. Presumably they arose from the earlier discussions with practicing professionals. The first hypothesis tested by the authors is that PJPA will be manifest as a series of actions and events. I suggest that the format of the requests for descriptions of events virtually guaranteed that the data would support this hypothesis. While the requests were open-ended with respect to the content of the description, the format the subjects were required to use was very structured. Subjects were asked to list the event that began the judgment process and the event that ended it. Then they were asked to list all the intervening events in chronological order. A numbered listing was provided for this purpose. Such a structure communicates to the subjects the type of process the question-writer was after. The resulting descriptions, therefore, may not reflect judgment processes, in general, but a subset of judgments which could be structured as a series of actions and events. Accordingly, given the instructions, the test of this hypothesis cannot be valid. When one requests descriptions of a series of events and actions, that is what one will get.

A second hypothesis tested in this section also seems marginal. The
authors hypothesize that the judgment situations reported not only will include auditing judgments, where auditing is narrowly defined as audit procedures, but also auditing, accounting, staffing, and client advice. The instructions requested that the subjects present a situation in which they were required to use their professional judgment. I do not think that anyone would have argued ex ante that all public accounting judgment is related to auditing procedures. Additionally, the subjects may have included tax- and management-consulting professionals who have no contact with judgments on auditing procedures. What might be of interest with respect to the different judgment contexts reported is whether the judgment processes differ by context. However, such an analysis is not presented by the authors. Prior professional accounting judgment research has dealt primarily with auditing tasks. In the interest of generalizing these results, knowledge of any judgment differences across tasks would have been helpful.

The third hypothesis the authors test is that PJPA is largely a process of responding to the environment. The test procedure involves classifying the reported initial event in all retrospective judgment descriptions as environment-initiated, subject-initiated, or environment-subject-interaction-initiated. The basis for these categorizations is not clear from the paper. For example, a phone call to the subject is an environmental event, while assignment of the subject to a task is an interaction event. At some level all human behavior can be considered a result of an interaction between the subject and the environment and, thus, I find the distinctions drawn here unclear at best. Gibbins (1984) suggests that all judgments are part of a continuous process, hence, even designating one event as the beginning of the judgment process is somewhat artificial.

Four additional hypotheses are tested with respect to the judgment descriptions. These hypotheses are derived from the cognitive learning-model propositions. A central feature of this model is the idea of judgment templates which are stored in long-term memory. Templates are part of a large associative memory complex, and the retrieval of a specific template is triggered when certain cues in the environment match the cues for a particular template. Associated with each judgment template is a response or judgment which becomes the subject’s response preference for the current judgment situation. Once this response preference is selected it must be considered with respect to the particular aspects of the current situation before it is undertaken. This later activity is known as the bridging process.

The first two hypotheses derived from this model are (1) that problem definition occurs relatively early in the judgment process
and (2) that problem definition precedes response preference. My understanding of the connection here is that template retrieval is equivalent to problem definition and, based on the model, template retrieval must precede (or coincide with) response preference. After completing their judgment descriptions, the subjects were asked to label the points at which the problem definition was completed and a response preference was identified. The authors report that, most of the time, problem definition either precedes or coincides with response preference. However, a disturbing aspect of the data is that problem definition follows response preference almost as often as it precedes it. This result could cast unfavorable light on the model, but the deficiency may not be related to the model. Specifically, the deficiency may be that the subjects' idea of problem definition is not synonymous with template retrieval. Problem definition may include part of the bridging process, that is, considering aspects of the particular situation in light of a general response preference. In that case, problem definition could be expected to follow response preference. Hence, the measure called problem definition really may be a noisy measure of template retrieval, and almost any result here could be considered consistent with the model.

A third hypothesis derived from the model is that a response preference will arise before the subject is able to justify it. Justification of a response to oneself and others is part of the bridging process. For most of the cases the subjects reported that the point at which a response could be justified followed (or coincided with) the point at which the response preference was identified. But what about the minority of cases for which the reverse was true? What does it mean to be able to justify a response that has not yet been considered as a choice? My theory here is that several possible responses may arise in connection with a template, and the subjects only reported the point at which the response that ultimately was chosen became the preferred one. A group of potential responses, all of which could have been justified, may have arisen earlier in the judgment process. In this light it would be interesting to consider whether the subjects reported responses as potential solutions which were later disregarded. My guess is that this was uncommon in the successful judgment descriptions. In reporting successful judgment processes one of the things that is left out is the blind alleys and the false starts.

The fourth hypothesis considered is that more experienced professionals develop response preferences earlier than less experienced professionals. This hypothesis is based on the idea that more well-developed, well-used templates allow for faster retrieval. This hy-
hypothesis is not supported, however, based on the correlation between number of years experience and the point at which a response preference was identified. Although the model may give a correct prediction on this point, the experience measure is noisy. I assume that accountants with different levels of experience reported different types of judgment situations. In fact, the authors claim that audit assistants reported most of the audit decisions while managers and partners reported most of the staffing decisions. In the first place, the more experienced auditor may have no more experience with the particular judgment contexts they report than do the less experienced auditors with their contexts. In the second place, different decision contexts, routine versus nonroutine in particular, may result in response preference identification at different points in the judgment process.

Next I wish to consider the statistical analyses used by Gibbins and Emby to describe their data and conduct tests of hypotheses. The use of any inferential statistic requires that one deal with a sample of observations which is somehow representative of a large population. The sample used here is a selection of practicing public accountants from large public accounting firms in four countries. The authors state that the sample is nonrandom with respect to the public accounting profession in general, and they make several qualifying statements about the nature of the population to which the results might apply. They also state that only simple tests of their hypotheses were conducted to avoid making assumptions about the nature of the underlying population. However, the use of any statistics at all requires assumptions to be made about the nature of the underlying population from which the representative sample is presumed to be drawn. Though the sample here is not random, I see no reason to believe it is not representative of the population of practicing public accountants in large public-accounting firms. If the authors are concerned that they have a biased sample, they could have compared characteristics of their sample with characteristics of a random sample or the population in general. In any case, the statistics assume that the sample represents a population.

The hypotheses tested in the study are presented in nonstandard statistical format. It is standard to present a null hypothesis, which one then attempts to reject in favor of an alternative hypothesis which is supported by theory or intuition. The authors present only their alternative hypotheses and report that the data support (or do not support) their hypotheses. Whether or not their evidence favors their hypotheses depends on the nature of the null hypothesis against
which the alternatives have been tested. Statistically, data are not used to support a particular hypothesis but rather to reject a null hypothesis at a particular level of confidence. Failure to reject a null hypothesis does not mean that the data do not support the alternative, but rather that the support is not so strong as to reject the null or that the statistical tests used are not powerful enough to reject the null. A careful presentation of the null and alternative hypotheses allows the reader to assess the appropriateness of the particular statistics used as well as the validity of the results, and, thus both should have been included in the Gibbins and Emby paper.

The statistical analysis of the close ended portion of the questionnaire involves a comparison of mean ratings across items (to assess relative importance), t-tests of differences between item means (to test for differences in response to different items), and interitem correlations. The random variable dealt with in these analyses is discrete with four response categories labeled 0, 1, 2, 3. The scaling of the variable is ordinal (rather than interval or ratio). Consequently, the use of the mean statistic is inappropriate. If the intervals of the scale were changed with the order unaffected, several of the mean comparisons reported in the paper would reverse direction. For example, if the scale values were 0, 15, 16, 30 instead of 0, 1, 2, 3, several of the reported comparisons would be altered.

Based on the mean statistics the authors conclude that the item with the highest mean is the most important to PJPA. Even if the mean were an appropriate measure, the conclusion that the item with the highest mean is the most important is inexact. Each subject who marked the highest-mean item as high could also have marked another item as high, which he/she felt was even more important. Alternatively, if subjects disagreed about which item was most important, an item that no one considered most important could receive the highest-mean score. The appropriate conclusion would have been that there was the highest agreement on the importance of this particular item.

The use of t-tests on the ordinal data also is inappropriate. The null hypothesis which the authors apparently wish to reject is that the distributions of responses are the same for each item. The t-test assumes continuous random variables from normal distributions (neither of which is true here) and tests for differences in the means of these distributions. As mentioned earlier, the mean is a noninformative statistic for ordinal data, and the failure of the other t-test assumptions to hold make these tests inappropriate here. All of the reported test results could be changed by changing the interval of the scaling. The
tests used should have been based on some statistic employing ranks or a multinomial distribution.

Little information is provided in the paper on the nature of the correlation statistic used for the interim correlations. Are they Pearson correlations (assuming the same continuous normal random variables as the earlier t-tests), or are they Spearman rank correlations which would have been more appropriate? In either case, the conclusions the authors draw from the correlation results are inappropriate given only tests of association and the absence of underlying theory. From the same type of reasoning and existing evidence on storks in England one could conclude that storks bring babies.

The statistic used to analyze the open-ended portion of the questionnaire (the judgment descriptions) is the percentage of the way through the judgment description at which a critical event (problem definition, response preference, and so on) occurred. These percentages were calculated based on the number of steps reported up to the event versus the total number of steps reported in the process. Again, the measures employed for the relative position of critical events in the process are just ordinal. Though this makes the sign tests that the authors use in the hypothesis testing appropriate, the percentage comparisons are not. The size of each step in each process may be far from uniform. Some steps may take more time than others and different individuals may report steps at varying degrees of aggregation. For example, is confirmation of accounts receivable one step or many? The rejection of the hypothesis that more experienced auditors develop response preferences more quickly may be partially a function of differential aggregation of events as well as the problems with the experience measure discussed earlier.

In conclusion, the weakest aspect of the study is the statistical analysis. Although the questionnaire could have been improved to allow for additional tests of reliability, the suggestions made represent minor improvements. Some trade-offs always must be made in questionnaire design. I think the study deals with an important issue and has collected some potentially interesting data. With some additional analysis the study can provide further insights and testable hypotheses on professional judgment in public accounting.

As an extension of the current study I suggest that the authors use their questionnaire on other groups of nonaccounting professionals. One important motivation for the study is a consideration of how the environment and decision requirements of public accounting cause the decision processes of accounting professionals to differ from those of other professionals. Knowledge of these differences could
allow us to evaluate results in the expert judgment literature in general, as to its applicability to the professional public accountant. While this study identifies environmental and decision related factors which are important to PJPA, it does not identify them as specific to PJPA or as general features of professional judgment. A replication of this study with other professional groups could make such discrimination possible.

REFERENCE

Evidence on the Nature of Professional Judgment in Public Accounting: Discussion

Professional judgment is an elusive concept. An analysis of its use in public accounting not only is a worthwhile academic effort as a contribution to the body of knowledge of accounting and auditing, but also is thought-provoking for the practitioner engaged in the judgment process.

Gibbins and Emby have based their analysis on the responses of public accounting professionals to two different questionnaires developed to elicit information regarding the characterization of professional judgment in public accounting (PJPA); identification of situations involving PJPA; identification of the sequence of events constituting the judgment process; and the role of consultation in the judgment process.

In reviewing the survey techniques employed, there could be some debate regarding population definition, sample size determination, the structure of the questionnaires, bias inherent in the questions, and so on. However, such debate could distract us from the subject of the analysis. Accordingly, I will not offer commentary on the survey itself, but will focus on the use of the results.
The authors have dealt with the use of judgment not only in audit activity, but have taken a broader view of public accounting to encompass judgment processes regarding accounting and reporting, personnel management, business risk, and client development. This approach provides a good representation of the practice of public accounting. An audit professional, particularly at the executive level, has many responsibilities which go beyond auditing. With these observations as background, let us now turn to an analysis of survey responses.

CHARACTERIZATION OF PJPA

In response to questions regarding the characterization of PJPA, the first group of respondents singled out justifiability as the most important quality. The second group most frequently preferred the following characterization statement: "I will try to make complete judgments that will prevent future problems, not just solve the present problem." These responses appear to be consistent. The first demonstrates a need for the results of PJPA to withstand the effects of subsequent reevaluation. The second demonstrates a need to withstand the effects of subsequent events. Both demonstrate a long-range view of the judgment process.

SITUATIONS INVOLVING PJPA

The authors collected information on judgment across a broad base of public accounting activities by nonrestrictive wording of the questionnaires. Of 118 situations identified by the respondents, only 32, or approximately 27 percent, related to auditing procedures. Very few of the audit situations were identified by senior personnel. Because the respondents, personnel in international accounting firms (presumed to be primarily audit personnel), most certainly devote a major portion of their professional time to audit activities, some interesting issues are raised here:

- What differences exist in the relative need for, and use of, PJPA in conducting audit versus other activities?
- Was there (in spite of efforts to screen out the effect) a response bias favoring nonaudit situations, relating to the level of interest in the situations identified or other factors?
- Does the low response of senior personnel regarding audit procedures appropriately reflect the extent of their expected involvement in audit activities?
These issues are not resolved in the analysis and may be worthy of further investigation.

Perhaps the most interesting facet of the responses is the overwhelming indication that the use of PJPA is a reactive process. Three of 118 situations identified represent activities initiated by the respondents. Before we conclude that the respondents lack initiative, however, it is appropriate to note that the 118 responses also included 68 events arising from interaction with the environment. Nevertheless, with nearly all of the identified situations being environmental or arising from interaction with the environment, the orientation of the profession to serving clients as their needs arise is clearly underscored.

**SEQUENCE OF EVENTS**

The authors developed a series of hypotheses regarding the sequence of events in PJPA that were tested using situations reported by the respondents and a hypothetical situation. Although minor variations were identified with respect to the timing and sequencing of events, in general, the hypotheses were supported. In particular, analysis of the data allowed the authors to conclude that preferences are developed rapidly upon definition of circumstances, and justification follows.

Worthy of our attention is the clear disparity between the respondents’ conceptual view of the critical elements of PJPA and their application of PJPA to their own experiences and a hypothetical event. As noted earlier, when one group of respondents was asked to identify the qualities or characteristics of PJPA, justifiability was singled out as most important. However, when this same group of respondents was asked to identify the most crucial elements of PJPA based on personal experience, in the aggregate, problem definition and results preferences were identified by a substantial margin as more crucial than justification. Less persuasive, but not inconsistent, results were obtained when the second group responded regarding a standard situation.

We may perhaps conclude, therefore, that while the respondents are concerned, at a conceptual level, with applying PJPA on a basis that can be subsequently justified and defended, the realities of the environment and specific circumstances require a reordering of the significance of the elements of PJPA. In practice, respondents may feel a need to place greater emphasis on problem definition and preferences created by environmental forces, for which justification may be more of a subsequent, than a concurrent, process.
As previously acknowledged, the inclusion of both audit and nonaudit activities in the survey is most appropriate in developing a realistic depiction of public accounting. However, when analyzing the sequence of events in the PJPA process, it seems essential that client service situations be segregated from all others and that the two groups be independently considered. This approach is necessary to acknowledge the significant impact of the client on the PJPA process, particularly with respect to preference development. In client service situations, problem definition and preference development may occur simultaneously at the initiation of the judgment process. In situations not related to client service, a more “normal” sequence of events may occur. Because the sequence data for client and nonclient situations has been commingled, the summary data presented cannot be used to analyze these situations distinctly.

The one hypothesis identified as unsupported predicted a positive correlation between respondent experience and speed of preference development. Based on my knowledge of numerous professionals engaged in practice, it is difficult to accept that such a correlation does not exist. In fact, this result is so surprising that it raises questions regarding the effectiveness of the survey in this area.

CONSULTATION

Perhaps the most clearly identified result of the authors' survey is the importance of consultation to PJPA. The results indicate that 93 percent of the respondents consulted at some point in connection with the situations reported. However, the respondents varied in their perceptions of the usefulness of consultation depending upon their rank and relative position to those consulted. In addition, there was variation in the use of consultation in the several phases of the PJPA process. Nonetheless, a strong preference was shown for consultation conducted between the points at which preference is established and action is taken.

All of the respondents were representatives of large international accounting firms. Based on my own experience in such a firm, frequent use of consultation seems quite natural. The structure of the firms, the human resources available within them, and the encouragement to consult embodied in the policies and procedures of many or all of the firms have resulted in frequent consultation becoming an important professional tool, and an identified element of quality control.
CONCLUSION

Gibbins and Emby have taken a realistic approach to their analysis of PJPA. They have attempted to study the use of judgment in all areas of public accounting activity, rather than auditing activity alone. As I have indicated, the results and interpretations are, for the most part, consistent with observations I have made over the course of more than a dozen years as an employee and a partner of a large international firm. I encourage additional study of PJPA with the hope it will provide a better understanding of this complex process.
The Demand for Quality-Differentiated Audit Services in an Agency-Cost Setting: An Empirical Investigation

Agency theory has been used recently to explain both a demand for audit services (for example, see Jensen and Meckling (1976) and Watts and Zimmerman (1983)), and the heterogeneous demand for audit quality (for example, see DeAngelo (1981)). Empirical evidence supporting the former type of demand has been reported by Chow (1982) who found agency-cost variables to be associated with the voluntary decision to engage external auditors in the pre-Securities Acts era. However, there presently is little empirical evidence linking agency-cost variables and the choice of quality-differentiated auditors.1

The objective of this study is to investigate the association between agency-cost variables and the use of quality-differentiated auditors. To do so, it is first necessary to identify the quality-differentiated suppliers of audit services. Although this is not an easy task, some prior research has reported that Big Eight auditors may be regarded as quality-differentiated (for example, see Arnett and Danos (1979), Nichols and Smith (1983), and Palmrose (1984)). Industry specialists

Editors' Note: This paper was anonymously refereed prior to acceptance.
(i.e., the largest suppliers of audit services within client industries) potentially are another group of quality-differentiated auditors (for example, see Palmrose (1984) and Shockley and Holt (1983)). Both Big Eight and industry specialist audit firms are utilized as quality-differentiated suppliers in this study. Consequently, the empirical tests are joint tests of (1) the association of agency cost variables with the use of quality-differentiated auditors, and (2) the Big Eight and industry specialist firms as quality-differentiated suppliers.

In examining the association between agency-cost variables and the use of quality-differentiated auditors, this study develops a model of auditor choice and tests the model using logit analysis. Each group (Big Eight/non-Big Eight and industry specialist/nonspecialist) is separately analyzed when the model is tested. The theoretical rationale supporting an agency-cost framework for the demand for quality-differentiated audit services is reviewed in the first section. Then agency cost variables and their anticipated effect on the choice of auditors are discussed in the second section. The remaining sections describe the data collection process, and the results of the study.

THEORETICAL RATIONALE: AGENCY COSTS AND THE DEMAND FOR QUALITY-DIFFERENTIATED AUDIT SERVICES

Agency theory helps explain the demand for audit services in terms of the role of audited information in the contractual relationships between organizational participants. Although there are multiple types of participants, investors, creditors, and managers tend to be the focus of agency relationships. An agency relationship is characterized as:

A contract under which one or more persons (the principal(s)) engage another person (the agent) to perform some services on their behalf which involves delegating some decision-making authority to the agent. (Jensen and Meckling 1976, p. 308)

From this perspective, an organization can be viewed as a nexus of contracts, both written and unwritten (Jensen and Meckling 1976), and contractual participants can be aligned into principal-agent relationships, such as investors (principals) and managers (agents).

Under the assumption of utility maximization "there is good reason to believe that the agent will not always act in the best interests of the principal" (Jensen and Meckling 1976, p. 308). The principal's welfare loss from this divergent behavior is one type of agency cost. However, it is generally recognized that the principal anticipates this divergent behavior of the agent and this expectation is incorporated in the contract price. Furthermore, it is generally recognized that
this cost can be reduced, thereby benefiting contractual participants, by alternate arrangements that induce the agent to behave as if he were maximizing the principal’s welfare. This objective can be accomplished by establishing appropriate incentives for the agent and by controlling the agent’s behavior through monitoring activities. Auditing is a form of monitoring activity, and the cost of audit services is another type of agency cost.

The incentive to engage auditors is enhanced by the information asymmetry problem inherent in the principal-agent relationships. Usually the agent has more inside information than he provides to the principal. In addition, the agent may have an incentive to bias the information supplied. The agent’s incentive and opportunity for exploiting the information asymmetry are assumed to be anticipated by the principal, and contracts are adjusted to reflect that expectation. Auditing helps to reduce the agents’ incentives and opportunities for exploiting the information asymmetry, and thereby reduce agency costs.\(^5\) In addition, differences in agents’ incentives and opportunities imply differences in the demand for audit service quality. Relationships with greater incentives and opportunities (that is, those with greater agency costs) may benefit (in terms of overall cost reduction) from the services of higher-quality auditors, where higher quality entails greater information credibility (less bias), for example, see Dopuch and Simunic (1982).\(^4\)

To summarize, in real world organizations, “agency problems arise because complete, fully contingent, costlessly enforceable contracts do not exist” (Klein 1983). Agency costs include

\[
\text{...the costs of structuring, monitoring, and bonding a set of contracts among agents with conflicting interests, plus the residual loss incurred because the cost of full enforcement of contracts exceeds the benefits. (Fama and Jensen 1983, p. 527)}
\]

Auditing is a form of monitoring and although it is costly, audited financial information is demanded because such information can help reduce the overall cost of contracting. Since the cost of auditing increases with audit service quality (for example, see Palmrose 1984), companies with greater costs of agency relationships are most likely to benefit from higher-quality audit services. The higher audit costs are justified by the expected greater reduction in other agency costs, so that overall agency costs would be lower. The next section describes the circumstances that might reflect differential agency costs and proposes variables to capture these differential costs in order to examine their association with the choice of quality-differentiated auditors.
A MODEL OF AUDITOR CHOICE

This section uses an agency cost framework to develop a model of auditor choice, that is, the likelihood a company uses a quality-differentiated auditor. The variables proposed for inclusion in the model capture differences in the costs of agency relationships. In each case, when the agency relationship involves higher costs (excluding auditing), the services of quality-differentiated auditors are likely to be of the greatest benefit.

Agency Cost Variables

Owner/Management Control

The implications of the separation of corporate ownership and control have long been debated. One aspect of this debate is summarized by Dhaliwal, Salamon, and Smith (1982) as follows:

...managers of large corporations with diffuse ownership (hereafter called management-controlled or MC firms) have considerable discretion in guiding the affairs of their firms and that this discretionary power is used to divert some resources from corporate shareholders. (p. 42)

In the context of agency theory, corporations that are controlled by managers (agents) rather than owners have greater potential wealth transfers (agency costs) than those controlled by owners.

One difficulty in formulating an owner/management control variable is defining the proportion of common stock ownership that constitutes control, since, in large public corporations, a stockholder with a relatively small percentage of shares may be in control. The literature has generally used 10 percent or more of the voting stock as a basis for defining control (for example, see Dhaliwal, Salamon, and Smith 1982). For the purposes of this study, a dichotomous variable is used in the model with 10 percent as the cut-off percentage. Companies in which one party owns 10 percent or more of the common stock are classified as owner-controlled (OC) and are assigned a value of 0; all others are classified as management-controlled (MC) and are assigned a value of 1.5

Leverage

Another type of contractual relationship with the potential for conflicts of interest is that of bondholders/shareholders. Smith and Warner (1979) outline four major sources of potential conflicts:

(1) Claim Dilution — If the firm sells bonds, and the bonds are priced assuming that no additional debt will be issued, the value of
the bondholders' claims is reduced if the firm issues additional debt of the same or higher priority.

(2) Dividend Payout — If a firm issues bonds and the bonds are priced assuming the firm will maintain its dividend policy, the value of the bonds falls if the firm raises the dividend rate and finances the increase by reducing investment. In the limit, if the firm sells all its assets and pays a liquidating dividend to the stockholders, the bondholders are left with worthless claims.

(3) Asset Substitution — If a firm sells bonds for the stated purpose of engaging in low variance projects and the bonds are valued at prices commensurate with that low risk, the value of the stockholders' equity rises, and the value of the bondholders' claim falls if the firm then substitutes projects with a higher variance rate for those with a low variance rate.

(4) Underinvestment — Myers (1977) suggests that a substantial portion of the value of the firm is composed of intangible assets in the form of future investment opportunities. A firm with outstanding bonds can have incentives to reject projects which have a positive net present value if the benefit from accepting the project accrues to the bondholders. (Pp. 118–19)

The greater the proportion of debt in a company's capital structure, the greater the potential for wealth transfers (that is, agency costs) from bondholders to shareholders. As Chow (1982) notes,

Researchers generally agree that, as the proportion of debt in a firm's capital structure increases, shareholders have a greater incentive to transfer wealth from the bondholders. This, in turn, implies a greater probability of suboptimal investment policies, and a greater potential gain to shareholders from contracting with bondholders (to limit the shareholders' ability to transfer wealth). (P. 275)

A leverage variable is included in the model to capture these potential agency cost differences. This variable is measured as the ratio of book value of long-term debt (excluding deferred credits, primarily deferred taxes) to total equity. Total equity is defined as the book value of long-term debt (excluding deferred credits), plus the book value of preferred stock, plus the market value of common stock (number of shares of common stock outstanding times the market value per share).

Size

Agency costs can be expected to vary with the size of the organization. As the size of the organization increases the number of agency
relationships also increases (for example, manager/shareholder, shareholder/bondholder, and top management/other levels of management). Increases in organization size also are likely to increase the remoteness of the principal(s) and, thus, decrease the ability of the principal(s) to observe the actions of agent(s). Finally, the likelihood is greater that the agent(s) has (have) a comparative advantage in producing financial information so that the attendant information asymmetry problems are increased with increases in organizational size. Consequently, the magnitude of the potential wealth transfers (agency costs) should increase with organizational size. Thus, a size variable, measured as the book value of total assets, is included in the model.

**Number of Subsidiaries**

The complexity and diversity (including geographical dispersion) of an organization's activities and operations can be expected to influence agency costs. As with size, complexity and diversity mean a potential for increases in the number of agency relationships as well as the remoteness of principals from the observation of agents' actions. Accordingly, the number of subsidiaries is included in the model to capture organizational diversity and complexity.

**Management Compensation Plan(s)**

Shareholders may contract with management under arrangements that attempt to mitigate the divergence of interests between the two groups. One such arrangement is to write compensation agreements tied to organizational performance. Accounting numbers can be utilized in these contracts (i.e., incorporated for defining appropriate performance goals). To alter effectively the manager's behavior (and thus, the shareholder's expectation of the magnitude of potential wealth transfers from divergent behavior), the financial information on which such contracts are based needs to be credible (for example, unbiased). One mechanism for increasing the credibility of financial information is to utilize the services of a quality-differentiated auditor.

One of the difficulties with including a variable in the model to reflect the existence of management compensation plans that are tied to accounting numbers is that full and complete information on such plans is not publicly available. As a proxy, the study uses a dichotomous variable based on the existence of a bonus or profit-sharing plan, tied to accounting numbers, that is publicly disclosed. In the model, the variable is 1 if such a plan is publicly disclosed, and 0 otherwise.
Exchange Listing

Finally, a variable is included in the model to capture some related aspects of auditor choice. This variable reflects the stock exchange listing of the company; it is denoted by 1 if the company is listed on the NYSE, and 0 otherwise. The dichotomous classification recognizes that NYSE companies are subject to particular types of monitoring arrangements as a condition of listing, including outside directors and audit committees, and these factors may affect the choice of auditors.

Summary: Model of Auditor Choice

The agency-cost variables are incorporated in a multiple regression model for logit analysis, as follows:

\[
AF = b_0 + b_1 OMC + b_2 D/E + b_3 \ln \text{assets} + b_4 \ln \text{subs} + b_5 \text{Mgt comp} + b_6 \text{Ex} + u,
\]

where

the dependent variable includes

\(AF\) = audit firm, the probability of a company using a quality-differentiated audit firm, with assigned values of (1) if Big Eight, (0) if non-Big Eight and (1) if industry specialist, (0) if nonspecialist;

the independent agency-cost variables include

\(OMC\) = owner/management controlled, (0) if one party owns 10 percent or more of the common stock (OC), (1) otherwise (MC);

\(D/E\) = leverage, ratio of book value of long-term debt (excluding deferred credits, primarily deferred taxes) to total equity; where total equity is the book value of long-term debt (excluding deferred credits), plus the book value of preferred stock, plus the market value of common stock;

\(\ln \text{assets}\) = book value of total assets (natural log form);

\(\ln \text{subs}\) = number of subsidiaries (natural log form);
Mgt comp = existence of a management compensation plan tied to accounting numbers, where (1) if bonus or profit-sharing plan tied to accounting numbers is publicly disclosed, (0) otherwise;  
Ex = exchange listing, where (1) NYSE, (0) otherwise; and  
u = error term.

All of the independent variables are expected to have positive coefficients. The various sources of information for measuring each of the variables are summarized in Table 1.

**DATA COLLECTION**

Data were gathered on companies in several industries, including office equipment, retail trade (food), electric utilities, gas utilities, and combination utilities. Although the utilities are regulated to a greater extent than office equipment and retail trade, no specific alterations are made in the model to consider that difference. Instead, the relevance of the variables included in the model may be considered to reflect regulatory effects, at least to some extent. To avoid the

<table>
<thead>
<tr>
<th>Variable</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audit firm</td>
<td>Annual financial report or 10-K</td>
</tr>
<tr>
<td>Ownership</td>
<td>Proxy statement</td>
</tr>
<tr>
<td>Leverage:</td>
<td></td>
</tr>
<tr>
<td>Long-term debt (BV)</td>
<td>Annual financial report</td>
</tr>
<tr>
<td>Preferred stock (BV)</td>
<td>Annual financial report</td>
</tr>
<tr>
<td>Common stock (number of shares outstanding)</td>
<td>Annual financial report</td>
</tr>
<tr>
<td>Market value of common stock</td>
<td><em>Wall Street Journal</em></td>
</tr>
<tr>
<td>Size:</td>
<td></td>
</tr>
<tr>
<td>Total assets</td>
<td>Annual financial report, 10-K, or <em>Standard &amp; Poor's Register of Corporations, Directors, and Executives</em></td>
</tr>
<tr>
<td>Number of subsidiaries</td>
<td>Annual financial report, 10-K, or proxy statement</td>
</tr>
<tr>
<td>Existence of management compensation plan tied to accounting numbers</td>
<td>Annual financial report or <em>Wall Street Journal</em></td>
</tr>
<tr>
<td>Exchange</td>
<td></td>
</tr>
</tbody>
</table>
potential of confounding differences among industries in the model variables, the data analysis is conducted separately for each industry.7

The final sample consists of companies meeting all of the following criteria: (1) Listed in Standard & Poor's Register for Corporations, Directors and Executives (1981) or Who Audits America (1980), (2) Not a foreign-based company, subsidiary, or conglomerate,∗ (3) Proxy statements, annual financial reports, and 10-Ks (for 1980 or 1981 fiscal years) available in either the University of California-Berkeley or Stanford University library.

The number of companies, per industry, included in the sample is presented in Table 2. The table also provides a breakdown of the number of companies which used Big Eight/non-Big Eight and industry specialist/nonspecialist audit firms. The industry specialist designation was based upon the audit firm industry market share information summarized in Table 3. In particular, the largest or two largest suppliers in each industry were designated industry specialists (see Palmrose 1984).

Descriptive statistics, by industry, are presented in Table 4 for the dependent and independent variables. Noteworthy information in this table includes the lack of any management compensation plans tied to accounting numbers in the utilities; the relatively high percentage of management-controlled utilities; the relatively high percentage of Big Eight firms in the sample (particularly in the utilities); and differences across industries in the average values of the variables.

### Table 2. Number of Companies in the Sample

<table>
<thead>
<tr>
<th>Industry</th>
<th>Total Companies</th>
<th>Big Eight*</th>
<th>Non-Big Eight</th>
<th>Industry Specialist</th>
<th>Non-specialist</th>
</tr>
</thead>
<tbody>
<tr>
<td>Office equipment</td>
<td>102</td>
<td>81</td>
<td>21</td>
<td>23</td>
<td>79</td>
</tr>
<tr>
<td>Retail trade</td>
<td>65</td>
<td>55</td>
<td>10</td>
<td>13</td>
<td>52</td>
</tr>
<tr>
<td>Electric utilities</td>
<td>39</td>
<td>36</td>
<td>3</td>
<td>21</td>
<td>18</td>
</tr>
<tr>
<td>Gas utilities</td>
<td>43</td>
<td>29</td>
<td>4</td>
<td>20</td>
<td>23</td>
</tr>
<tr>
<td>Combination utilities</td>
<td>27</td>
<td>26</td>
<td>1</td>
<td>21</td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td>276</td>
<td>237</td>
<td>39</td>
<td>98</td>
<td>178</td>
</tr>
</tbody>
</table>

* Big Eight firms are defined as the following: Arthur Andersen & Co. (AA), Arthur Young (AY), Coopers & Lybrand (CL), Deloitte Haskins & Sells (DHS), Ernst & Whinney (EW), Peat, Marwick, Mitchell & Co. (PMM), Price Waterhouse & Co. (PW), and Touche Ross (TR).
A comparison of group means (Big Eight/non-Big Eight and industry specialist/non-specialist) for each of the independent variables is provided in Table 3. The comparisons reveal some directional support for the association of the agency cost variables with the use of Big Eight auditors. With the exception of debt/equity in office equipment, retail trade, and gas utilities, companies using Big Eight auditors tend to have higher mean values for each of the agency cost variables. There are only two other exceptions to this pattern.

Table 3. Audit Firm Industry Market Shares

<table>
<thead>
<tr>
<th>SEC-SIC Codes</th>
<th>AA</th>
<th>AY</th>
<th>CL</th>
<th>DHS</th>
<th>EW</th>
<th>PMM</th>
<th>PW</th>
<th>TR</th>
<th>Non-Big Eight</th>
</tr>
</thead>
<tbody>
<tr>
<td>357 Office equipment</td>
<td>.03</td>
<td>.08</td>
<td>.05</td>
<td>.09</td>
<td>.01</td>
<td>.17</td>
<td>.54</td>
<td>.01</td>
<td>.02</td>
</tr>
<tr>
<td>541 Retail trade</td>
<td>.05</td>
<td>.01</td>
<td>.13</td>
<td>.08</td>
<td>.06</td>
<td>.34</td>
<td>.08</td>
<td>.17</td>
<td>.08</td>
</tr>
<tr>
<td>491 Electric utilities</td>
<td>.34</td>
<td>.00</td>
<td>.17</td>
<td>.33</td>
<td>.00</td>
<td>.03</td>
<td>.11</td>
<td>.00</td>
<td>.02</td>
</tr>
<tr>
<td>492 Gas utilities</td>
<td>.49</td>
<td>.06</td>
<td>.00</td>
<td>.13</td>
<td>.01</td>
<td>.12</td>
<td>.05</td>
<td>.08</td>
<td>.06</td>
</tr>
<tr>
<td>493 Combination utilities</td>
<td>.25</td>
<td>.03</td>
<td>.15</td>
<td>.34</td>
<td>—</td>
<td>.01</td>
<td>.21</td>
<td>—</td>
<td>.01</td>
</tr>
</tbody>
</table>

.00 indicates audit firm industry market share less than .01
— indicates audit firm had no clients in industry
* industry specialist
Sources of data: WAA (1980) and SPR (1981) with computations based on client revenues.

Table 4. Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>Office Equipment</th>
<th>Retail Trade</th>
<th>Electric and Combo Utilities</th>
<th>Gas Utilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total assets: mean</td>
<td>$ 741MM</td>
<td>$ 338MM</td>
<td>$ 2,351MM</td>
<td>$ 651MM</td>
</tr>
<tr>
<td>(st dev)</td>
<td>(3,129MM)</td>
<td>(574MM)</td>
<td>(2,455MM)</td>
<td>(1,014MM)</td>
</tr>
<tr>
<td>Debt+total equity: mean</td>
<td>.20</td>
<td>.32</td>
<td>.56</td>
<td>.45</td>
</tr>
<tr>
<td>(st dev)</td>
<td>(.25)</td>
<td>(.25)</td>
<td>(.07)</td>
<td>(.14)</td>
</tr>
<tr>
<td>Subsidiaries (#): mean</td>
<td>6.3</td>
<td>8.2</td>
<td>1.9</td>
<td>6.3</td>
</tr>
<tr>
<td>% mgt control</td>
<td>.41</td>
<td>.28</td>
<td>.92</td>
<td>.79</td>
</tr>
<tr>
<td>% mgt comp plan tied to accg numbers</td>
<td>.13</td>
<td>.22</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>% NYSE</td>
<td>.36</td>
<td>.34</td>
<td>.86</td>
<td>.51</td>
</tr>
<tr>
<td>% Big Eight auditors</td>
<td>.79</td>
<td>.85</td>
<td>.94</td>
<td>.91</td>
</tr>
<tr>
<td>% industry specialists</td>
<td>.23</td>
<td>.20</td>
<td>.64</td>
<td>.47</td>
</tr>
</tbody>
</table>
Table 5. Comparison of Group Means

<table>
<thead>
<tr>
<th>Office Equipment</th>
<th>Big Eight</th>
<th>Non-Big Eight</th>
<th>Industry Specialist</th>
<th>Non-specialist</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assets mean (66/88MM)</td>
<td>$315 (158)</td>
<td>$315 (158)</td>
<td>$315 (158)</td>
<td>$315 (158)</td>
</tr>
<tr>
<td>(2.98/3.18)</td>
<td>(3.72/3.75)</td>
<td>(3.72/3.75)</td>
<td>(3.72/3.75)</td>
<td>(3.72/3.75)</td>
</tr>
<tr>
<td>Debt/equity (3.8/4.8)</td>
<td>1.38</td>
<td>1.38</td>
<td>1.38</td>
<td>1.38</td>
</tr>
<tr>
<td># subs</td>
<td>7.2</td>
<td>7.2</td>
<td>7.2</td>
<td>7.2</td>
</tr>
<tr>
<td>% mgmt control</td>
<td>43</td>
<td>43</td>
<td>43</td>
<td>43</td>
</tr>
<tr>
<td>% mgmt comp plan</td>
<td>43</td>
<td>43</td>
<td>43</td>
<td>43</td>
</tr>
<tr>
<td>NYSE (310)</td>
<td>(310)</td>
<td>(310)</td>
<td>(310)</td>
<td>(310)</td>
</tr>
<tr>
<td>Retail Trade</td>
<td>3.92*</td>
<td>3.92*</td>
<td>3.92*</td>
<td>3.92*</td>
</tr>
<tr>
<td>Assets mean (61/81/MM)</td>
<td>$315 (215)</td>
<td>$315 (215)</td>
<td>$315 (215)</td>
<td>$315 (215)</td>
</tr>
<tr>
<td>(3.03/3.21)</td>
<td>(3.47/3.73)</td>
<td>(3.47/3.73)</td>
<td>(3.47/3.73)</td>
<td>(3.47/3.73)</td>
</tr>
<tr>
<td>Debt/equity (3.7/4.7)</td>
<td>1.48</td>
<td>1.48</td>
<td>1.48</td>
<td>1.48</td>
</tr>
<tr>
<td># subs</td>
<td>4.7</td>
<td>4.7</td>
<td>4.7</td>
<td>4.7</td>
</tr>
<tr>
<td>% mgmt control</td>
<td>43</td>
<td>43</td>
<td>43</td>
<td>43</td>
</tr>
<tr>
<td>% mgmt comp plan</td>
<td>43</td>
<td>43</td>
<td>43</td>
<td>43</td>
</tr>
<tr>
<td>NYSE (282)</td>
<td>(282)</td>
<td>(282)</td>
<td>(282)</td>
<td>(282)</td>
</tr>
</tbody>
</table>

This directional support for the use of Big Eight firms, the management compensation plan variable in office equipment, and the subsidiaries variable in electric and combination utilities. Asset differences are statistically significant (SL ≤ .05) in three of the industries; and differences in debt/equity, subsidiaries, and percentage of NYSE listings are significant in two industries.

The comparisons for industry specialists/nonspecialists are less supportive of the association between agency-cost variables and the use of quality-differentiated auditors. A possible exception is office equipment companies; those using industry specialists have higher
Although the univariate tests have an advantage in that measure-
ment errors in one independent variable do not affect the test
results for other variables (see Chow, 1982), any interrelationship among the
mean values for all variables except debt/equity, and the differences
in both the percentages of the management costs and the use of industry
specialists are statistically significant. In these industries, only the
difference in subsidiaries of gas utilities is statistically significant.

### Electric and Combination Utilities

<table>
<thead>
<tr>
<th></th>
<th>Big Eight</th>
<th>Non-Big Eight</th>
<th>t-value</th>
<th>Industry Specialist</th>
<th>Non-specialist</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assets: mean</td>
<td>$2,486MM</td>
<td>$2,257MM</td>
<td>6.44*</td>
<td>$2,446MM</td>
<td>$2,184MM</td>
<td>.42</td>
</tr>
<tr>
<td>(std dev)</td>
<td>(2,472MM)</td>
<td>(2,900MM)</td>
<td></td>
<td>(2,630MM)</td>
<td>(2,158MM)</td>
<td></td>
</tr>
<tr>
<td>Debt/equity</td>
<td>.565</td>
<td>.536</td>
<td>.84</td>
<td>.563</td>
<td>.554</td>
<td>-.09</td>
</tr>
<tr>
<td>(std dev)</td>
<td>(.057)</td>
<td>(.089)</td>
<td></td>
<td>(.067)</td>
<td>(.072)</td>
<td></td>
</tr>
<tr>
<td># subs</td>
<td>1.9</td>
<td>2.8</td>
<td>-.29</td>
<td>1.8</td>
<td>2.0</td>
<td>-.19</td>
</tr>
<tr>
<td>(std dev)</td>
<td>(2.7)</td>
<td>(2.6)</td>
<td></td>
<td>(2.8)</td>
<td>(2.5)</td>
<td></td>
</tr>
<tr>
<td>%mgt control</td>
<td>.94</td>
<td>.75</td>
<td>.74</td>
<td>.91</td>
<td>.96</td>
<td>-.87</td>
</tr>
<tr>
<td>(std dev)</td>
<td>(.25)</td>
<td>(.50)</td>
<td></td>
<td>(.30)</td>
<td>(.20)</td>
<td></td>
</tr>
<tr>
<td>%NYSE</td>
<td>.89</td>
<td>.50</td>
<td>1.33</td>
<td>.86</td>
<td>.88</td>
<td>-.20</td>
</tr>
<tr>
<td>(std dev)</td>
<td>(.32)</td>
<td>(.58)</td>
<td></td>
<td>(.35)</td>
<td>(.34)</td>
<td></td>
</tr>
</tbody>
</table>

### Gas Utilities

| Assets: mean           | 685MM     | 327MM         | .67     | 814MM               | 510MM          | .95     |
| (std dev)              | (1,055MM) | (390MM)       |         | (1,230MM)           | (783MM)        |         |
| Debt/equity            | .449      | .472          | -.18    | .451                | .452           | -.02    |
| (std dev)              | (.134)    | (.295)        |         | (.148)              | (.145)         |         |
| # subs                 | 6.7       | 2.8           | 2.64*   | 9.1                 | 4.0            | 2.27*   |
| (std dev)              | (7.6)     | (1.7)         |         | (9.3)               | (4.2)          |         |
| %mgt control           | .80       | .75           | .21     | .85                 | .74            | .88     |
| (std dev)              | (.41)     | (.50)         |         | (.37)               | (.45)          |         |
| %NYSE                  | .51       | .50           | .05     | .50                 | .52            | -.14    |
| (std dev)              | (.51)     | (.58)         |         | (.51)               | (.51)          |         |

* SL < .05
independent variables may diminish the usefulness of these comparisons. Table 6 presents a matrix of simple correlation coefficients for the variables in the model. From this table it appears that the size variable is significantly (SL ≤ .05) correlated with several other variables, particularly the subsidiaries and exchange variables.

**Regression Results**

Because of the dichotomous dependent variable and some dichot-

**Table 6. Matrix of Simple Correlation Coefficients**

<table>
<thead>
<tr>
<th></th>
<th>O/M Control</th>
<th>Debt/Equity</th>
<th>Ln Assets</th>
<th>Ln Subs</th>
<th>Mgt Comp</th>
<th>Exchange</th>
<th>B8/NB8</th>
<th>IS/NS</th>
</tr>
</thead>
<tbody>
<tr>
<td>O/M Control:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OE</td>
<td>1.00</td>
<td>-.04</td>
<td>.27*</td>
<td>.22*</td>
<td>.16</td>
<td>.36*</td>
<td>.13</td>
<td>.22*</td>
</tr>
<tr>
<td>RT</td>
<td>.05</td>
<td></td>
<td>.21</td>
<td>.04</td>
<td>-.15</td>
<td>.28*</td>
<td>.07</td>
<td>-.14</td>
</tr>
<tr>
<td>ECU</td>
<td>.03</td>
<td></td>
<td>.32*</td>
<td>.20</td>
<td>.22</td>
<td>.17</td>
<td>-.10</td>
<td></td>
</tr>
<tr>
<td>GU</td>
<td>-.12</td>
<td>.24</td>
<td>-.07</td>
<td></td>
<td></td>
<td>.18</td>
<td>.03</td>
<td>.14</td>
</tr>
<tr>
<td>Debt/equity</td>
<td>1.00</td>
<td>-.23*</td>
<td>-.09</td>
<td>-.01</td>
<td>-.05</td>
<td>-.36*</td>
<td>-.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-.04</td>
<td>.15</td>
<td>-.04</td>
<td>-.05</td>
<td>-.27*</td>
<td>-.11</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.10</td>
<td>.24*</td>
<td>.04</td>
<td>.10</td>
<td>-.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-.13</td>
<td>-.28</td>
<td>-.08</td>
<td>-.05</td>
<td>-.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ln assets</td>
<td>1.00</td>
<td>.52*</td>
<td>.18</td>
<td>.71*</td>
<td>.40*</td>
<td>.41*</td>
<td>.17</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.42*</td>
<td>.15</td>
<td>.66*</td>
<td>.41*</td>
<td>.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.34*</td>
<td>.68*</td>
<td>.45*</td>
<td>.97</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.64*</td>
<td>.77*</td>
<td>.07</td>
<td>.14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ln subs</td>
<td>1.00</td>
<td>.16</td>
<td>.45*</td>
<td>.22*</td>
<td>.20</td>
<td>.08</td>
<td>-.02</td>
<td>.15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-.12</td>
<td>.42*</td>
<td>.08</td>
<td>.01</td>
<td>.20</td>
<td>-.08</td>
<td>-.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.44*</td>
<td>.14</td>
<td>.35*</td>
<td></td>
<td>.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mgt comp</td>
<td>1.00</td>
<td>.08</td>
<td>-.02</td>
<td>.15</td>
<td>.14</td>
<td>.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exchange</td>
<td>1.00</td>
<td>.28*</td>
<td>.47*</td>
<td>.31*</td>
<td>.05</td>
<td>.27*</td>
<td>-.03</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>.34*</td>
<td>.02</td>
<td></td>
<td>.34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B8/NB8</td>
<td>1.00</td>
<td>.27*</td>
<td></td>
<td></td>
<td></td>
<td>.21</td>
<td>.34*</td>
<td>.30</td>
</tr>
<tr>
<td>IS/NS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
</tr>
</tbody>
</table>

1OE: office equipment; RT: retail trade; ECU: electric and combination utilities; GU: gas utilities.
* SL ≤ .05
omous independent variables, the multivariate analysis is conducted using logit. The industry regression results are reported in Table 7 for the Big Eight/non-Big Eight dependent variable and in Table 8 for the industry specialist/non specialist dependent variable. The significance of the auditor choice model can be assessed using the model chi-square statistic which is twice the difference in the log likelihood of the reported model from the likelihood based on the intercept only. Three of the four Big Eight/non-Big Eight regressions (office equipment, retail trade, and electric and combination utilities) are significant (SL <= .05). Only office equipment is significant for the industry specialist/non specialist regressions, however. The tables also provide the D-statistic for each regression. The D-statistic can be interpreted in a similar manner as the R-square under standard

Table 7. Regression Results

<table>
<thead>
<tr>
<th></th>
<th>Office Equipment (n=102)</th>
<th>Retail Trade (n=65)</th>
<th>Electric and Combo Utilities (n=56)</th>
<th>Gas Utilities (n=43)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept coefficient</td>
<td>-7.331</td>
<td>-16.370*</td>
<td>-51.993*</td>
<td>358</td>
</tr>
<tr>
<td>Owner/mgt control</td>
<td>.538</td>
<td>.256</td>
<td>-1.407</td>
<td>.498</td>
</tr>
<tr>
<td></td>
<td>(.672)</td>
<td>(1.124)</td>
<td>(2.530)</td>
<td>(1.411)</td>
</tr>
<tr>
<td>Debt + total equity</td>
<td>-2.310*</td>
<td>-2.894</td>
<td>6.313</td>
<td>.082</td>
</tr>
<tr>
<td></td>
<td>(1.156)</td>
<td>(1.544)</td>
<td>(8.665)</td>
<td>(4.178)</td>
</tr>
<tr>
<td>Ln assets</td>
<td>.529*</td>
<td>1.075*</td>
<td>2.872*</td>
<td>.060</td>
</tr>
<tr>
<td></td>
<td>(.256)</td>
<td>(.485)</td>
<td>(1.399)</td>
<td>(.646)</td>
</tr>
<tr>
<td>Ln subsidiaries</td>
<td>.074</td>
<td>-.487</td>
<td>-2.720</td>
<td>.622</td>
</tr>
<tr>
<td></td>
<td>(.422)</td>
<td>(.551)</td>
<td>(1.396)</td>
<td>(.720)</td>
</tr>
<tr>
<td>Mgt compensation</td>
<td>-1.237</td>
<td>.699</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.882)</td>
<td>(1.347)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exchange</td>
<td>.287</td>
<td>6.894</td>
<td>-2.607</td>
<td>-.651</td>
</tr>
<tr>
<td></td>
<td>(1.040)</td>
<td>(16.384)</td>
<td>(3.177)</td>
<td>(1.679)</td>
</tr>
<tr>
<td>$X^2_{of}$ (significance)</td>
<td>25.63*</td>
<td>20.44*</td>
<td>18.98*</td>
<td>1.22</td>
</tr>
<tr>
<td></td>
<td>(.0002)</td>
<td>(.002)</td>
<td>(.0019)</td>
<td>(.9432)</td>
</tr>
<tr>
<td>D statistic</td>
<td>.22</td>
<td>.26</td>
<td>.24</td>
<td>.03</td>
</tr>
<tr>
<td>Rank correlation</td>
<td>.62</td>
<td>.75</td>
<td>.96</td>
<td>.33</td>
</tr>
<tr>
<td>Fraction of concordant pairs</td>
<td>.80</td>
<td>.87</td>
<td>.98</td>
<td>.64</td>
</tr>
</tbody>
</table>

* SL <= .05
Table 8. Regression Results

<table>
<thead>
<tr>
<th></th>
<th>Office Equipment (n=102)</th>
<th>Retail Trade (n=65)</th>
<th>Electric and Combo Utilities (n=66)</th>
<th>Gas Utilities (n=43)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept coefficient</td>
<td>-5.814</td>
<td>-8.349</td>
<td>-4.676</td>
<td>-2.232</td>
</tr>
<tr>
<td>(standard error)</td>
<td>(3.218)</td>
<td>(6.136)</td>
<td>(5.152)</td>
<td>(7.255)</td>
</tr>
<tr>
<td>Owner/mgt control</td>
<td>.187</td>
<td>-.1140</td>
<td>-1.181</td>
<td>1.618</td>
</tr>
<tr>
<td></td>
<td>(.603)</td>
<td>(.933)</td>
<td>(1.255)</td>
<td>(.989)</td>
</tr>
<tr>
<td>Debt/total equity</td>
<td>-.545</td>
<td>-1.078</td>
<td>-.477</td>
<td>2.967</td>
</tr>
<tr>
<td></td>
<td>(1.539)</td>
<td>(1.579)</td>
<td>(4.056)</td>
<td>(2.770)</td>
</tr>
<tr>
<td>Ln assets</td>
<td>.201</td>
<td>.395</td>
<td>.360</td>
<td>-.198</td>
</tr>
<tr>
<td></td>
<td>(.187)</td>
<td>(.342)</td>
<td>(.270)</td>
<td>(.435)</td>
</tr>
<tr>
<td>Ln subsidiaries</td>
<td>-.175</td>
<td>-.072</td>
<td>-.259</td>
<td>1.147*</td>
</tr>
<tr>
<td></td>
<td>(.271)</td>
<td>(.313)</td>
<td>(.383)</td>
<td>(.593)</td>
</tr>
<tr>
<td>Mgt compensation</td>
<td>.777</td>
<td>.732</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.752)</td>
<td>(.777)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exchange</td>
<td>1.920*</td>
<td>-.016</td>
<td>-.873</td>
<td>-1.202</td>
</tr>
<tr>
<td></td>
<td>(.801)</td>
<td>(.975)</td>
<td>(1.049)</td>
<td>(1.071)</td>
</tr>
<tr>
<td>X^2 df</td>
<td>25.40*</td>
<td>6.40*</td>
<td>2.66*</td>
<td>10.93</td>
</tr>
<tr>
<td>(significance)</td>
<td>(.0003)</td>
<td>(.3801)</td>
<td>(.7526)</td>
<td>(.0528)</td>
</tr>
<tr>
<td>D statistic</td>
<td>.21</td>
<td>.10</td>
<td>.04</td>
<td>.23</td>
</tr>
<tr>
<td>Rank correlation</td>
<td>.61</td>
<td>.40</td>
<td>.20</td>
<td>.49</td>
</tr>
<tr>
<td>Fraction of concordant pairs</td>
<td>.80</td>
<td>.69</td>
<td>.59</td>
<td>.74</td>
</tr>
</tbody>
</table>

* SL<=.05

(ordinary least squares) regressions. For additional information, Tables 7 and 8 disclose the fraction of concordant pairs' predicted probabilities and responses and the rank correlation between predicted probabilities and responses. For a null hypothesis of no association, tests of significance evaluate deviations from 50 percent of the percentage of concordant pairs, and zero of the rank correlation.

The coefficients of the control variable (owner/management-controlled organization) and the compensation plan variable (existence of a management compensation plan tied to accounting numbers) are positive in a majority of the regressions but not significant in any. Coefficients for the debt/equity, Ln subsidiaries, and exchange variables have signs opposite from what was expected (that is, negative) in five of the eight regressions. Each of these coefficients is significant.
in one regression. For In subsidiaries and exchange, the coefficients are positive when significant, while the debt/equity coefficient is negative when significant. The size variable coefficient is positive in all Big Eight/non-Big Eight regressions and significant in three. The size-variable coefficient is not significant in any of the industry specialist/nonspecialist regressions, although it is positive in all but one.¹⁰

**Analyses to Control for the Effects of Size**

The matrix of simple correlation coefficients (Table 6) reveals that some of the variables in the model are significantly associated with the size variable. Because the regression results could be influenced by these interrelationships, additional analyses were conducted.

One approach to the problem of multicollinearity, in the context of multiple regression analysis, is to drop one of the interrelated variables from the analysis. This approach was followed by dropping the size variable from each of the logistic regressions. The results, however, are not much different from those reported in Tables 7 and 8. In terms of significant differences, only the exchange coefficient in the Big Eight/non-Big Eight electric and combination utilities regression were altered (that is, to a positive and significant coefficient with the size variable deleted). In addition, the industry specialist/nonspecialist regression for gas utilities was significant (SL ≤ .05) with the assets variable deleted.¹¹

Another approach to control for the effects of company size in the Big Eight/non-Big Eight grouping of auditors also was utilized. Specifically, a matched-pairs procedure was employed, in which each company in the sample with a non-Big Eight auditor was matched with a similar-sized company in the same industry having a Big Eight auditor. Size was based on total assets and "similar size" was defined as total assets within $10 million. This matching procedure produced 35 pairs, including 21 in office equipment, 9 in retail trade, 2 in electric and combination utilities, and 3 in gas utilities.

Multivariate analysis was performed for this matched-pairs sample using the logit model on the differences in the independent variables (see Bowen, Noreen, and Lacey 1981; and Kelly 1984).¹² In the logit regression, the coefficients for the owner/management control, debt/equity, and subsidiaries variables were negative; and, the coefficients for the management compensation and exchange variables were positive. None of the coefficients was significant and the model was not significant (SL ≤ .05).
Analysis of Companies Which Changed Auditors

It could be argued that the tests of the model of auditor choice could be strengthened if the data set were restricted to companies clearly involved in auditor-choice decisions as revealed by changes in auditors. Since little is known of the magnitude of the cost of changing auditors, it is possible that such costs are a sufficient impediment to using what would otherwise be the "desired" auditor from the standpoint of current agency costs. Limiting the sample to companies involved in changes in auditors helps to mitigate this problem.

To carry out this analysis, companies in the sample that changed auditors within five years prior to, or two years subsequent to, the financial information fiscal year end were identified using Who Audits America (1980). Only office equipment produced an adequate sample size — 32 observations with the dependent variable determined as follows:

1 = Change from Big Eight to Big Eight
   Change from non-Big Eight to Big Eight (n=15)
   Change from non-Big Eight to Big Eight (n= 8),

0 = Change from Big Eight to non-Big Eight
   Change from non-Big Eight to non-Big Eight (n= 5)
   Change from non-Big Eight to non-Big Eight (n= 4),

1 = Change from industry specialist to industry specialist (n= 0)
   Change from nonspecialist to industry specialist (n= 6),

0 = Change from industry specialist to nonspecialist (n= 2)
   Change from nonspecialist to nonspecialist (n=24).

The logit regression results for this sample subset are reported in Table 9. Although the model is significant for both the Big Eight/ non-Big Eight and industry specialist/nonspecialist regressions, nothing appears to be gained in terms of statistically significant coefficients within the model, as compared with the results for office equipment as a whole. Additional regressions were run deleting the size variable because of concerns that the reported results might be affected by any interrelationship between size and other variables in the model. None of the remaining variable coefficients were significant when the size variable was deleted.

DISCUSSION OF RESULTS

Although the univariate tests tend to be directionally supportive of the association of the agency cost variables with the use of Big Eight auditors, the results for the Big Eight/non-Big Eight logit
Table 9. Regression Results for Office Equipment Companies That Changed Auditors

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Big Eight/Non-Big Eight</th>
<th>Industry Specialist/Nonspecialist</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept coefficient</td>
<td>-23.835*</td>
<td>4.059</td>
</tr>
<tr>
<td>(standard error)</td>
<td>(10.732)</td>
<td>(13.347)</td>
</tr>
<tr>
<td>Owner/mgt control</td>
<td>.351</td>
<td>.044</td>
</tr>
<tr>
<td></td>
<td>(1.323)</td>
<td>(1.522)</td>
</tr>
<tr>
<td>Debt+total equity</td>
<td>-2.172</td>
<td>-4.155</td>
</tr>
<tr>
<td></td>
<td>(2.144)</td>
<td>(4.283)</td>
</tr>
<tr>
<td>Ln assets</td>
<td>1.563*</td>
<td>-.307</td>
</tr>
<tr>
<td></td>
<td>(.675)</td>
<td>(.836)</td>
</tr>
<tr>
<td>Ln subsidiaries</td>
<td>-.862</td>
<td>-3.086</td>
</tr>
<tr>
<td></td>
<td>(.864)</td>
<td>(2.088)</td>
</tr>
<tr>
<td>Mgt compensation</td>
<td>6.282</td>
<td>5.659</td>
</tr>
<tr>
<td></td>
<td>(18.033)</td>
<td>(3.163)</td>
</tr>
<tr>
<td>Exchange</td>
<td>4.503</td>
<td>7.422</td>
</tr>
<tr>
<td></td>
<td>(22.967)</td>
<td>(4.745)</td>
</tr>
<tr>
<td>$X_{50}$</td>
<td>17.36*</td>
<td>15.46*</td>
</tr>
<tr>
<td>(significance)</td>
<td>(.0081)</td>
<td>(.0170)</td>
</tr>
<tr>
<td>D statistic</td>
<td>.41</td>
<td>.38</td>
</tr>
<tr>
<td>Rank correlation</td>
<td>.82</td>
<td>.83</td>
</tr>
<tr>
<td>Fraction of concordant pairs</td>
<td>.91</td>
<td>.92</td>
</tr>
</tbody>
</table>

* SL<.05

Note: In both regressions n=32.

Regressions are not particularly enlightening. Almost all of the Big Eight/non-Big Eight regressions are significant, but only the size variable is significant. This result, that is, larger companies are more likely to use Big Eight auditors, already is widely known. Furthermore, this result is somewhat unsatisfactory because size can proxy for other factors that may influence auditor choice but are unrelated to the demand for quality-differentiated services. For example, non-Big Eight firms may be less likely to meet the terms and conditions (such as adequate staff sizes and locations) that are prerequisites for auditing larger companies.

There are several potential reasons why the other variables failed to contribute much in the way of explanation in the Big Eight/non-
Big Eight regressions. First, the difficulties in measuring agency-cost variables probably weakened the analysis. Such difficulties are evident in the specification of the model of auditor choice and reflect problems with the public availability of information and the nature of the guidelines imposed for using that information.

In addition, the auditor choice model is developed from a demand perspective. It assumes, therefore, that auditors are willing to supply services to any client. In reality, this is unlikely to be the case. Therefore, the auditor choice/client acceptance decision is more complex than implied by the model. Particularly for quality-differentiated auditors, audit risk and the exposure to reputation diminishment are assessed before taking on a client. To some extent, the results for the debt/equity variable may reflect these considerations. Recall that the debt/equity coefficient is generally negative and is negative and significant for office equipment. This result can be interpreted as suggesting that companies with higher debt/equity ratios may represent greater risk and, thus, may not be cost effective clients for quality-differentiated auditors.

With the exception of office equipment, the industry specialist/non-specialist results from the univariate and regression analyses are generally not supportive of the association of agency-cost variables with the use of quality-differentiated auditors. These results may not only reflect weaknesses noted in the discussion of the Big Eight/non-Big Eight results, but the results may indicate a lack of support for the industry specialists as quality-differentiated auditors.15

SUMMARY AND CONCLUSIONS

This study has developed and tested a model of auditor choice to examine the association between agency-cost variables and the use of quality-differentiated auditors. Two classifications of quality-differentiated auditors were utilized, Big Eight/non-Big Eight and industry specialists/non-specialists. The model was tested on a sample of companies from several industries, including office equipment, retail trade (food), electric/combination utilities, and gas utilities.

For all but one variable (leverage), the univariate tests tended to be directionally supportive of the association between the agency cost variables and the use of Big Eight auditors. The multivariate tests were conducted using logit analysis. For the Big Eight/non-Big Eight, the logit regressions were generally significant, with company size (total assets) being the primary explanatory variable. The generally weak results for the other agency-cost variables was attributed to several factors, including difficulties in defining and measuring the
variables. Furthermore, the Big Eight/non-Big Eight results (particularly with respect to the leverage variable) seemed to suggest the necessity for recognizing additional complexities in the supply/demand for quality-differentiated audit services.

With respect to the industry specialists/nonspecialists, office equipment was the only industry that provided even weak evidence in either the univariate tests or the logit regressions of an association between the agency-cost variables and the use of industry specialists. Given the joint nature of the underlying tests (see the first section), these results may indicate weaknesses in the classification of industry specialists as quality-differentiated suppliers rather than a lack of support for the agency-cost variables and the use of quality-differentiated auditors.

APPENDIX

To aid in comparing the results of this study with other related work, some summary information is provided as follows:

(1) Chow (1982) used logit analysis to assess the probability of a company being externally audited in 1926 (that is, a dependent variable set equal to 1 if the firm had an external auditor and 0 otherwise), with the following general results:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient Sign</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size (MV of OE + BV of debt)</td>
<td>plus</td>
</tr>
<tr>
<td>Debt/equity (BV of debt + size)</td>
<td>plus*</td>
</tr>
<tr>
<td>No. accdg measures in debt cov agreements (adjusted for the effect of leverage)</td>
<td>plus*</td>
</tr>
<tr>
<td>Exchange (1 if NYSE, 0 otherwise)</td>
<td>plus*</td>
</tr>
<tr>
<td>Owner/management control (measurement problems because of the lack of available data)</td>
<td>na</td>
</tr>
</tbody>
</table>

* Significance level = .05

(2) Eichenscher and Shields's (1984) study uses logit analysis and 1980 financial information to examine the probability of using a Big Eight auditor for a sample of AMEX companies, those that changed/did not change auditors between 1973–1978, with the following preliminary results:

<table>
<thead>
<tr>
<th>Coefficient Sign</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change</td>
</tr>
</tbody>
</table>

| Size (consolidated assets where 1 if $44 million or more, and 0 otherwise) | plus | plus | plus |
Leverage (total debt (excluding deferred credits) ÷ total assets where 1 if ≤ .53 and 0 otherwise) plus* minus minus

Control (% of voting stock owned by officers and directors where 1 if ≤ .26, and 0 otherwise) plus* minus plus

Audit committee (yes = 1, and 0 otherwise) plus* plus plus*

* Significance level = .05

(3) Healy and Lys's (1983) study uses logit analysis to examine the probability that a client will stay with an acquiring (Big Eight) auditor or switch back to a non-Big Eight auditor at the time of a Big Eight/non-Big Eight merger. The model includes several variables for expected changes in capital structure and growth, with the following preliminary results:

| Size (\ln BV of total assets) | plus* |
| Leverage (long-term liabilities ÷ BV total assets) | minus |
| Growth (average % growth in assets in 3 years prior to merger) | plus* |
| Change in debt (change in long-term liabilities in 3 years after merger as a % of total BV) | plus |
| Change in contributed capital (change in contributed capital during 3 years after merger ÷ beginning BV) | minus |

* Significance level = .05

(4) Simunic and Stein's (1984) study examining the use of Big Eight/non-Big Eight auditors with original issuances of securities to the public presents a preliminary result from univariate tests as follows:

Leverage (long-term debt ÷ total assets after new public issuance of securities) Average leverage ratio for companies using Big Eight auditor is significantly greater than those using non-Big Eight auditor

Ownership (ratio of common stock held by management, officers, and directors to total number of shares of common stock) Average ratio of (insider) ownership is greater for companies using non-Big Eight auditors, but not significantly greater

The latter three studies focus on the association between agency cost variables and the choice of Big Eight/non-Big Eight auditors. Although comparing the results of the studies is hampered to some extent by the differences in the models and the measurement of the variables, the results
provide mixed support for the agency-cost variable. Together the three studies include some information on size, leverage, and ownership, but the direction and significance of the association of these variables with the choice of Big Eight auditors varies across the studies.

NOTES

*The research assistance of Hay Young Chung and Changyong Ham and
the financial assistance of the University of California-Berkeley Professional
Accounting Program are gratefully acknowledged. This study has benefited
from the suggestions of John Eichenseer, Marcia Niles, and an anonymous
referee.

1. There are several studies that provide some preliminary evidence on
the association between agency cost variables and the use of quality-
differentiated auditors, that is, higher quality suppliers. These studies focus
on the choice of Big Eight or non-Big Eight auditors in conjunction with
the following events: a change in auditors (Eichenseer and Shields 1984),
audit firm mergers (Healy and Lys 1983), and new issuances of public
securities (Simunic and Stein 1984). See the Appendix for a further discussion
of these studies.

2. The largest suppliers are based on audit firm industry market shares.
In Shockley and Holt’s (1983) study, audit firm market share in the banking
industry helped to explain the differences in bankers’ perceptions of audit
firms. The authors speculated that the bankers viewed the audit firms with
the largest market shares in banking as having greater industry expertise
than other firms. Palmrose (1984) expanded on the rationale for the largest
auditors in an industry being quality-differentiated suppliers and tested the
hypothesis using audit fee data. Though the results were directionally
consistent with industry specialists as quality-differentiated suppliers, the
results were not statistically significant (SL = .05). This study provides an
alternate to the use of audit fees for gaining insight into industry specialists
as quality-differentiated suppliers.

3. In addition, the demand for the agent’s services may increase with
the agreement to provide audited financial information. As Beaver (1981,
p. 49) notes, “in the limit . . . the principal may be unwilling to enter into
a contract with the agent.”

4. The assumption that Big Eight and industry specialists are higher-
quality suppliers of audit services is akin to a maintained hypothesis. Several
authors provide theoretical rationales for why these suppliers’ services are
more credible or higher quality, including DeAngelo (1985), Dopuch and
Simunic (1982), Nichols and Smith (1983), Palmrose (1984), and Simunic

5. For most of the companies in the sample, the party owning 10 percent
or more of the common stock was on the board of directors and/or part
of management.

6. For example, the utility companies in the sample do not have
management compensation plans tied to accounting numbers (see Table 4).

7. Because of the relatively small number of observations for combi-
nation utilities and the similarity of operations and operating environments
for electric and combination utilities, these two groups are combined in
subsequent analyses.
8. For utilities with Big Eight auditors, the sample represents a subset (approximately 50 percent) of the companies meeting the first two criteria.

9. The fraction of concordant pairs indicates the percentage of all pairs of companies having different auditors where the higher of the two predicted probabilities is in conjunction with the quality-differentiated (that is, Big Eight or industry specialist) auditor.

10. The Appendix presents a summary of some related work, including Chow (1982), Eichenseher and Shields (1984), Healy and Lys (1983), and Simunic and Stein (1984) as a basis for comparing the results of this study.

11. It should be noted that this discussion of the deletion of the assets variable from the logit regressions ignores the potential omitted variable specification error.

12. In the multivariate analysis, each pair represented both a non-Big Eight and a Big Eight auditor, so that the dependent variable was no longer dichotomous. Instead, each pair was randomly assigned to one of the two auditor types and the direction of the differences in the independent variables was computed accordingly. For this analysis, the exchange variable was defined as NYSE = 3.0, AMSE = 2.0, and OTC = 1.0. The intercept was suppressed in running the logit regression.

13. In addition, the results are somewhat sensitive to the particular audit firms used as quality-differentiated suppliers. For example, the regression results for retail trade are strengthened when Touche Ross is added as an industry specialist. While the regression is not significant at the 5 percent level, it is at the 10 percent level. Also, the subsidiaries coefficient is significant (SL ≤ .05) and the compensation coefficient is nearly significant (SL ≤ .08).

REFERENCES


The Demand for Quality-Differentiated Audit Services in an Agency-Cost Setting: An Empirical Investigation: Discussion

In her study, Palmrose uses agency theory to develop hypotheses underlying the demand for higher-quality audit services. This seems to be a reasonable path to pursue, given that it has become popular to rely on agency theory to explain the general demand for auditing services. Not surprisingly, however, she encounters problems similar to those in extant audit agency theory research. These problems primarily involve the choice of variables and interpretation of results.

The specific hypotheses tested by Palmrose are that agency variables can explain the demand for Big Eight vs. non-Big Eight audit services and the demand for industry vs. nonindustry specialists. Both dichotomies are assumed to reflect differential qualities of audit services. Agency arguments are used to predict that those client firms which face higher “monitoring” costs will seek out the Big Eight and specialist audit firms because of their higher quality audit services.

While it is popular to use agency theory to explain the demand for auditing services, the arguments are frequently so vague that
what we obtain is really no more than rationalizations of the existence of auditing as an industry. In fact, I do not know of any formal theories which demonstrate that auditing is the optimal response to agency (monitoring) costs existing in manager-owner relationships. Ex post, we see that auditing survives in a competitive setting; therefore, auditing can be viewed as cost-effective. But agency theory does not demonstrate this; empirical facts do.

Empirical facts also force us to recognize that the auditing industry is composed of different kinds of auditing firms, some being very large with national reputations and some quite small with only local reputations. The former are assumed to provide higher-quality audit services and this is the main issue addressed in Palmrose’s study. But if agency theory has not yet successfully demonstrated how auditing represents the optimal response to agency costs arising between managers and owners in a general context, we must be skeptical about claims that it would provide a definitive basis for explaining the demand for higher-quality audits in particular.

Any study which attempts to test agency hypotheses relating to the differential quality of audits will necessarily be subject to criticisms of alternative hypotheses, alternative test variables, and the like. For example, the demand for Big Eight and specialist audit firms may reflect the fact that such firms can serve large clients in a more economical manner (economies-of-scale arguments). Alternatively, concentrations of markets by the Big Eight segment and by specialists could reflect monopolistic practices — a favorite accusation of legislators. In evaluating studies of the auditing industry we need to keep in mind that the lack of data on revenue and cost functions of auditing firms prevents us from distinguishing between these alternative theories. Hence, such studies must be viewed as tentative, with the selected hypotheses serving more to structure the collection and analysis of empirical data rather than to test specific theories. To illustrate, early in her paper, Palmrose states that “The variables proposed for inclusion in the [her] model capture differences in the costs of agency relationships...” One set of variables may be slightly more appealing than another set, but none will completely dominate all the others. So, in discussing her study, I shall accept her candidate variables in those terms.

**SELECTION OF AGENCY VARIABLES**

Let me proceed now to a discussion of Palmrose’s agency variables and of some problems encountered in operationalizing them. Her
selection was guided by her interpretation of agency arguments applied to the demand for higher-quality audit services and the fact that similar or equivalent variables have been associated with agency hypotheses in other agency studies.

Owner-Management

The first variable she discusses is one designed to distinguish between owner- and manager-controlled firms. The typical assumption in agency studies is that owner-controlled (that is, owner-managed) firms are less susceptible to agency problems of moral hazard and the like. This is certainly a hard variable to dismiss, particularly if we contrast the two possible extremes of control. At one end, we would have firms which are completely owned by the persons managing the firm. At the other, the managers would have a zero ownership interest in their firms. In the first case, the owner-managers would have no incentives to transfer wealth between owners/managers — they are the same persons. In the second, there is little doubt that moral hazard and related agency problems would be present.

Unfortunately, empirical studies can rarely use extreme cases like the foregoing in testing agency hypotheses. Data needed to test such hypotheses are available mostly for publicly traded firms which are of a much more varied mixture of ownership than the two cases cited. Practically all traded firms have managers with some ownership interest, but almost none is completely owned by the management. Empirical studies usually adopt an arbitrary rule to separate the two classes of firms, such as classifying firms as owner-controlled if the management owns 10 percent or more of the voting stock. However, 10 percent is an arbitrary cut-off, and researchers can obtain varied classifications using other measures of owner- vs. manager-controlled firms; for example, see Niehaus (1984). Hence, while few would argue that agency problems would be less severe for highly owner-controlled firms, empirical realities force us to use more arbitrary and less clear-cut measures of the two classes of firms.

Leverage

Of all the variables tested in studies of agency arguments, the one which seems to be the most consistently significant across the studies is leverage. The existence of debt raises the possibility of transferring wealth from creditors to owners/managers. To protect themselves, creditors insert debt covenants which make it illegal for owners/managers to engage in certain transactions (for example, to pay
dividends in the absence of sufficient earnings coverage). Leverage has been found to be a significant variable in studies of accounting choices, income smoothing, voluntary audits, and so on. On the basis of prior evidence alone, Palmrose is justified in including a leverage variable in her set of agency variables.

In support of this variable she argues why creditors would be happy to see management hire higher-quality auditors, but not why managers would be motivated to follow their desires. In order to argue that leverage explains the demand for higher-quality auditing services, Palmrose needs to show that catering to creditors’ desires increases the value of ownership interests (for example, their equities). This is possible, but the arguments were omitted in the version of the paper presented at this conference.

I might also note here that leverage is measured using book values of long-term debt (exclusive of deferred credits). But, book values may not be good measures of debt during periods of high interest rates such as those (the early 1980s) studied by Palmrose. Unfortunately, however, significant portions of a firm’s long-term debt could be privately placed, making it difficult to obtain market values for this kind of study.

Size and Number of Subsidiaries

Palmrose’s third and fourth variables were selected to reflect the increase in agency costs arising from more complex and numerous agency contracts in large, highly diversified firms. However, in reading the discussion supporting their selection, it seems that size and number of subsidiaries create more problems for top management than they do for the principals (that is, the owners) of firms.

I should also note here that size and number of subsidiaries, particularly if the latter are geographically dispersed, are good candidates for explaining why firms might select large national firms which have the expertise to handle diverse divisions. Thus, any significant coefficients observed for these variables would be equally consistent with hypotheses of economies of scale, as well as agency cost hypotheses.

Management Compensation

Her fifth variable is a dummy variable to reflect whether firms have compensation contracts with managers which are tied to accounting numbers. Certainly, the existence of such contracts can lead to potential agency problems. But if the contracts are based on GAAP,
it is not clear that having higher-quality audits will provide benefits that exceed the attendant extra costs.

More important, I am disturbed by the use of a dummy variable to capture the impact of management compensation contracts on agency costs. Researchers often comment on the problems encountered in trying to obtain data on the exact details of managerial compensation contracts that would more directly allow them to test the impact of compensation packages on managerial decisions. These problems are difficult to overcome. Nevertheless, the use of a dummy variable to reflect all of the many facets of managerial compensation contracts is surely a crude way to incorporate their potential impact in agency studies. Is there any compensation contract that does not depend on performance reports, whether directly or indirectly? Even contracts tied solely to stock price behavior will be influenced by accounting reports of earnings and financial positions of firms. I would wager that if we knew the details of the compensation contracts of all of the firms in Palmrose's samples, we would code all of them with a one for this dummy variable. The degree of reliance may be much less for regulated firms (for example, the utilities in her samples), but there would be sufficient reliance to code even those firms as having compensation contracts tied to accounting numbers. In short, I would not expect a significant coefficient for this agency variable simply because of the high probability of coding errors.

**Exchange Listing**

The last variable used in Palmrose's model is another dummy variable to indicate whether a firm is traded on the NYSE or elsewhere. The argument here is that NYSE firms are subject to additional monitoring requirements which may encourage them to use higher-quality audit firms. However, one could argue the other way, that the additional monitoring requirements reduce the need to rely on external auditors. Be that as it may, I would like to have seen some comparison of differential monitoring requirements before selecting such a variable for inclusion in the study. One statistical problem induced by including the exchange listing variable is that NYSE firms are generally larger than firms listed on other exchanges, so that this variable will be correlated with the size variable already mentioned.

**Omitted Variables**

Recall that I raised the question about whether auditing was an optimal response to agency problems. Surely, there are other forms
of monitoring which could be substituted for external audits — or at least used as complements. Firms may use active audit committees and formal internal audit divisions as monitoring mechanisms and these may reduce their demands for higher-quality auditing services. Similarly, creditors may be able to place their own representatives on the boards of firms that have high debt-equity ratios, relying on direct monitoring rather than external auditors. There probably are other types of monitoring mechanisms that can serve as substitutes or complements to external audits, but I do not recall any study of agency costs that tries to identify what these might be. The omission of these other monitoring variables will not, however, bias Palmrose's test results in favor of her hypotheses; rather their exclusion will make it more difficult to obtain positive results. Since the results of their tests were disappointing, she might consider whether omitted variables are a potential explanation of the results.

EMPIRICAL TESTS

Tests of the significance of Palmrose's agency variables were conducted at both a univariate and a multivariate level. The univariate tests were basically tests of differences between the means of the agency variables calculated for those clients that used Big Eight and specialist audit firms — those demanding higher-quality audits — and those that used non-Big Eight and nonspecialist firms. The multivariate tests were conducted using the logit form of regression model. Logit provides probabilities that firms will fall into one of the two classes — Big Eight or non-Big Eight in one set of regressions and industry specialist or nonspecialist in the other set. Logit has certain advantages when the dependent variable is a dichotomous — zero/one — variable, as would be the case for both sets of regressions.

The tests of the variables are cross-sectional tests using data taken from firms' 1980 or 1981 accounting reports and 10Ks. The firms selected were drawn from the office equipment, retail trade, electric, gas, and combination-utility industries. I could not find any arguments for the selection of firms from only two nonregulated industries, and why utilities were selected at all — except for the statement that the latter formed a basis for comparing the explanatory power of the model(s) across both regulated and nonregulated industries. However, the fact that utilities are regulated suggests an entirely different kind of monitoring (agency) problem than exists with nonregulated firms. Since this kind of research is still in its development stage, I would recommend addressing the problem of modeling the demand for
higher-quality audit services in at least two stages: (1) using about a
dozen or so different nonregulated industries and (2) using a modified
model to test other variables for regulated industries.

The sample selection procedures resulted in considerable degrees
of imbalance between Big Eight and non-Big Eight auditors. There
was also an imbalance between specialists and nonspecialists, but to
a lesser degree. Out of a total of 276 firms, 237 (almost 86 percent)
were audited by Big Eight firms. We need to keep this in mind, since
if we were to predict all firms were audited by Big Eight auditors,
we would only miss on 14 percent of them. The imbalance between
Big Eight and non-Big Eight is especially extreme with the utilities.
Out of 109 utilities, only 8 were audited by non-Big Eight auditors.
Consequently, I would not deem it worthwhile even to consider the
Big Eight vs. non-Big Eight issue with the utilities.

Univariate Tests

Descriptive statistics on the means of the test variables for the
various industries are provided in Palmrose's Table 4. Incidentally,
she combined the electric and "combination" utilities so the subse-
quent tests are run for four industries, rather than five. A more
informative table for our purposes is her Table 5 which provides the
means comparisons on an industry-by-industry basis according to the
two main issues of Big Eight vs. non-Big Eight and specialists vs.
nonspecialists.

The one consistent result shown there is that Big Eight clients are
larger than non-Big Eight clients for each of the four industries. A
similar result holds for the specialist clients relative to the nonspecialist
clients, although it is not as striking. Out of 44 tests of means
differences, only 12 were significant. Five of these 12 occurred with
size variables (assets and number of subsidiaries). What is missing
from the table, however, is a calculation of the number of significant
differences we would expect to observe by chance alone. My guess
is that 12 out of 44 would not be much different from chance results.

Logit Tests

Given the results shown in Table 5, I would have been surprised
if the logit regressions would produce significant coefficients on any
agency variables other than size. This expectation generally is con-
firmed in Table 7. The size coefficient is significant in three out of
the four industry regressions (the exception is gas utilities). The only
other variable achieving significance is the debt/equity variable in
the office equipment industry. However, it has the wrong sign relative to the agency arguments provided in the first section of Palmrose’s paper. The regression results for the specialist vs. nonspecialist issue are generally insignificant, with the only exception being the exchange variable coefficient for office equipment. These results appear in Table 8.

**SOME EXPERIMENTAL DESIGN ISSUES — POSSIBLE EXTENSIONS**

Let me comment next on some aspects of the experimental design of this study, and in the process offer some suggestions for extending it. As I have since learned, some of my suggestions have already been incorporated in the other studies referenced in Palmrose’s paper. If these studies obtain positive results, my suggestions will be vindicated. Otherwise, I will adopt the typical discussant’s ploy that the suggestions were at least worth a try.

When I first read her paper, I was struck by Palmrose’s decision to arbitrarily choose a year of financial results and then compare these in terms of Big Eight and non-Big Eight clients, and specialist and nonspecialist clients. It would seem that a more logical approach would have been to select firms and years in which decisions to select auditors were actually made. This would suggest a design which would focus on the levels of, or changes in, agency variables during the year of the auditor decisions. A control over size could then be incorporated by using a matched-pair design, selecting firms of similar sizes which did not make auditor decisions.

Palmrose made a pass at this approach by running her logit analysis on firms that switched auditors either within five years prior to the year in which the financial data were collected or two years subsequent. The implication is that she continued to use these firms’ financial data, collected in 1980–81, and merely reran the regressions on a special subclass of her sample. What I suggest is that data be collected for the years surrounding auditor decisions — say, prior to the year of the decision itself and one year subsequent. Assessing the changes in agency variables over these years, relative to changes in the same variables for a control group, would provide a more direct test of the agency arguments advanced in the paper. If the decision-firms were matched on size with control-firms, as she has since done in an extension of her study, she could avoid confounding due to economies of scale arguments in the selection of auditors. I might note here that once she controls for size the agency variables do not achieve
any significance, nor does her overall model. However, I believe these
results, again, merely reflect problems encountered in arbitrarily
selecting a year in which to collect the data, and perhaps the fact
that the sample sizes used in the matched pairs were small.*

In order to expand the sample, I would recommend that Palmrose
select firms from several different nonregulated industries and analyze
these separately. One possibility would be to select industries based
on a priori arguments about which would be more susceptible to
agency costs. For example, accounting issues might be crucial in some
industries, thereby creating the potential for higher monitoring costs.
I have no specific industries in mind here, but the possibility deserves
consideration.

My final comments deal with the use of logit and the manner in
which Palmrose presented the results in the conference version of
her paper. Logit (and its companion, probit) are relatively new
statistical tools used in accounting — at least relative to ordinary
least-squares. As such the outputs of regression-runs using logit need
more explanation than she provided. For example, I wonder how
many of the conference participants know the implications of the \( \chi^2 \),
the D, the rank correlation, and fraction of concordant pairs statistics
which appear at the bottom of Tables 7–9. Some details of their
implications would have been helpful.

Somewhat similarly, I believe a discussion of the test statistics used
to assess the significance of the coefficients in logit would have been
useful. My understanding is that we can employ the traditional \( t \)-
statistic to assess the significance of logit coefficients, but since the
distribution of the standard errors is known only asymptotically, we
must be suspicious of any marginally significant coefficients. This is
not a problem in the current study, since few coefficients obtained
significance. However, it may be in an extension using the matched-
pairs design already suggested.

Finally, the results of logit models should be tested against holdout
samples, either actual or simulated ones, using, say, the jackknife
procedures. As noted earlier, the logit regression model allows one
to predict the membership of observations in a probabilistic manner —
for example, we can predict whether a particular firm has a high
probability of being audited by a Big Eight or specialist firm. The
benchmark of comparison is the proportion of firms audited by Big
Eight or specialist firms. In Palmrose’s sample, the Big Eight audited
over 85 percent of the client firms. If the logit model has validity it
should allow one to exceed an 85 percent classification accuracy rate
on her sample, and better than the population proportion of Big
Eight audits on hold-out samples. In these terms, the regression models for the utility firms in her study would have had to have been close to perfect to outpredict the simple decision rule that all utilities would be audited by Big Eight auditors. A matched-pairs design would change the sample proportion to 50 percent, which would give logit models a better chance to outperform simple decision rules.

Of course, in using matched-pairs designs the researcher will face the common criticism that he/she stacked the sample in favor of positive results. A more defensible procedure is to select samples that reflect the actual population proportions of the variable of interest — for example, select Big Eight and non-Big Eight clients in proportion to what we actually observe in the real world. But this is the same kind of criticism that was levied against prior studies which examined the predictive ability of financial information in the area of bankruptcy. After a number of studies demonstrated that financial information could be used to predict bankruptcy at a higher level of accuracy than chance, researchers began to explore the ability to do so, using more representative sample proportions of failed vs. nonfailed firms (for example, see Zmijewski (1984)).

Tests of agency arguments within the context of auditing issues have not yet reached that level of achievement. Since these tests are based on relatively vague expressions of agency theories of auditing contracts, the researcher presently needs all the help available in the test designs. Accordingly, samples used to test the arguments should be chosen to increase the probability of achieving positive results. Once this is done we can then move to more representative sample designs as we attempt to replicate and extend our initial results. The main caveat at this stage is that researchers continue to recognize the tentative nature of their results until the results have been verified in subsequent studies.

NOTE

* To get some idea of how Palmrose's continuous variables might change over time, I had an assistant calculate the averages of asset size and debt/equity ratios for a sample of utility and office equipment firms for the years 1970, 1973, and 1977. I tried to obtain a sample of Big Eight and non-Big Eight clients in each industry, but like Palmrose I could not find many utilities which did not use a Big Eight auditor.

As we would guess, the average asset size for all the firms increased over this span of years. The electric utility averages went from $590 million to $113 billion. The average size of the gas utilities doubled during the same span of time. Whether asset sizes of non-Big Eight utility firms had the same growth is not known because all of these utilities were Big Eight clients.
Of more relevance, the average asset size of the office equipment firms which used Big Eight auditors increased from $700 million to $1.2 billion. In contrast, the average size of those using non-Big Eight audit firms increased from $84 million to $192 million. Recall that the logs of these numbers would be inserted in Palmrose's models, so whether the results would vary over time is not obvious.

What I found the most interesting were the changes in the debt/equity ratios over this seven-year span. The average debt/equity ratios of the electric utility firms declined from .473 to .42 (about two standard deviations), which exceeds the difference in the average ratios of the Big Eight and non-Big Eight electric utilities using Palmrose's numbers. An even more dramatic change occurred over the seven-year span for the sample of gas utilities selected by my assistant. The average of their debt/equity ratios went from .53 to .36, which was about a four-standard-deviation change in the means.

Finally, the changes in the average debt/equity ratios for the office equipment firms using Big Eight auditors went from .321 to .212, whereas those using non-Big Eight auditors had a change of only from .197 to .203. If Palmrose had chosen 1970 for the year of analysis she would likely have observed differences in the means of the debt/equity ratios for the office equipment firms that were consistent with her theoretical arguments, but not so in 1977 (and 1980).

REFERENCES

The Demand for Quality-Differentiated Audit Services in an Agency-Cost Setting: An Empirical Investigation: Discussion

RESEARCH OBJECTIVE

Palmrose's stated objective is "to investigate the association between agency-cost variables and the use of quality-differentiated auditors." To understand the paper, I found myself supplying the following unstated, but implicit, hypotheses:

$H_0$: Higher agency costs are associated with demands for higher audit quality; and

$H_1$: Higher agency costs are not associated with demands for higher audit quality.

An empirical test designed to discriminate between these two hypotheses could come out with any of three results measured in terms of the model coefficients:

Positive: this result would be satisfying, in that it would be consistent with the null hypothesis and the agency-theory prediction; however,
it would not prove the agency-theory prediction — it would merely fail to disconfirm it.

Negative: this unexpected result would refute the agency-theory prediction and would therefore have high surprise value and information content.

Insignificant: this result would leave us where we started.

The results of Palmrose's test were insignificant. Following are comments on some possible reasons for this.

WHAT IS A QUALITY-DIFFERENTIATED AUDITOR?

The key to this paper rests on the definition of "audit service quality," but this term is not defined by Palmrose. To understand the independent variable in her model and select relevant indicators, we need to understand its meaning. Following are three possible meanings to illustrate that the term is not unambiguous:

Alternative (1): Better audit quality means lower bias and variance of accounting errors, given a standard fineness of accounting reports. (This alternative is the one adopted in the professional standards and the one I accept. It may be approximately equivalently rephrased as a higher level of assurance that there are no material errors in the financial statements.)

Alternative (2): Better audit quality means greater auditor wealth to pay off claims for information error. (This alternative, which is inconsistent with professional standards, may be inferred from litigation claims.)

Alternative (3): Better audit quality means the client firm is less likely to fail or otherwise lead to an adverse investment outcome. (Though this alternative is also inconsistent with professional standards, it may be considered because of its reported incidence among users of financial statements.)

A further possibility not explicitly considered by Palmrose is that a single audit firm may deliver differential audit quality depending upon, for example, client size. As I have reported elsewhere, Peat Marwick uses an audit-planning materiality ("gauge") that is an exponential function of auditee size. This results in a relatively tighter precision for large auditees than for small ones. In a certain sense, it could be stated that Peat Marwick delivers higher quality for larger auditees. (I would argue that other firms do so also: first, if Peat Marwick were out of line, there would be an observable realignment of client relationships; second, a yet unreported AICPA
study by Carl Warren found a similar exponential relationship across approximately 60 audit firms.) Although Palmrose has a size variable in her model, it is supposed to indicate the level of agency costs. If, however, it is also indicative of differential quality, the two effects may be confounded.

HOW WOULD A USER DIFFERENTIATE AUDIT QUALITY?

A key to Palmrose’s argument is that users are able to differentiate higher-quality auditing so that they know how to demand it. Audits typically have an identical signal (the standard auditor’s report), except for the auditor’s name. (The possibility of various report modifications — such as qualified reports, adverse reports, or disclaimers — is ignored because they are rare in practice.) Therefore, the sole signal on which the user can differentiate is the auditor’s name.

Assume that the earlier presented first alternative for audit quality is accepted. Then better-quality auditing implies less error in the financial statements. However, there is no consistent feedback to users concerning which auditors are associated with less financial statement error.

How might users assess the quality of auditing in this feedback-poor setting? One possibility is that users form an impression about an audit firm’s quality from publicly available information about the firm. However, the most common and dramatic information about a firm comes from news reports about professional litigation against it. Users might, arguably, consider that the firm with the least litigation (relative to its size) provides the highest-quality audits.

There are several flaws, however, in this line of reasoning:

(1) The litigation signal is rare (which gives it high-surprise value and therefore high-recall value for users.) For SEC registrants (those with the highest propensity for litigation because of larger average size and various investor litigation advantages incorporated in the 1933 and 1934 Securities Acts), the probability of litigation is approximately .002 per year. Such a rare event cannot reliably discriminate audit quality.

(2) Litigation alleging audit failure arises in virtually all cases in which the auditee fails and in virtually no other. Owners and creditors who lost money typically proceed against the auditor as virtually the only “deep pocket” available after the failure. The correlation of accounting error and firm failure has not been studied (and, in fact,
the data to study it would be unavailable), but my opinion, based on reviewing a number of such cases, is that the correlation is negligible.

(3) Many lawsuits lack merit. In support of this assertion, it will be observed that many of the cases filed against auditors are either won outright by the auditors or settled for nuisance amounts relative to the original claims.

The lack of public information from which users can discriminate audit quality should have been dealt with explicitly by the author, as the ability of users to make such discriminations is essential to her hypotheses.

**IS BIG EIGHT OR INDUSTRY SPECIALIST A MEASURE OF AUDIT QUALITY?**

One possibility is that there are no significant differences in quality among firms regularly performing audits (many CPA firms perform few if any audits; it might reasonably be assumed that their quality is lower and that their work is perceived accordingly). Lack of quality differentiation might be the result of the well-known competitive market for audits in combination with professional standards setting an audit-quality floor and the peer-review program enforcing the minimum standards. Thus, audit quality could arguably have been driven into a range too narrow to be discriminated by Palmrose’s tests. Alternatively, there could be significant differences in audit quality, but Palmrose’s dependent variable might be misspecified (the chosen 0-1 partition might not maximize interclass variance).

**IS THE PREFERENCE FOR BIG EIGHT OR INDUSTRY SPECIALIST BASED ON QUALITY OR SOME OTHER ATTRIBUTE?**

There are possible explanations for selection of Big Eight or industry-specialist auditors other than audit quality:

(1) A large or geographically spread-out client requires an auditor of sufficient size or geographic spread to perform the engagement.

(2) A client has need for specialized tax or management advisory services that only the selected firm can supply.

(3) A client has been requested by a banker or underwriter to select a Big Eight firm. (This might be explained by a preference for high audit quality, but it might alternatively be explained by the banker’s or underwriter’s preference to avoid the costs of becoming familiar with a new auditor’s quality to discharge its “due-diligence” obligations.)
(4) A client may select a Big Eight auditor to maximize the depth of the pocket in case of later claims.

Thus, the author should consider whether the observed preferences reflect a demand for higher-quality auditing or some other attribute.

DO THE SELECTED INDEPENDENT VARIABLES MEASURE AGENCY COSTS?

Dopuch has discussed this matter thoroughly in his comments, but I agree that Palmrose's independent variables make sense, though their scaling seems rather coarse. However, I do not understand the author's expectation of a positive coefficient for the NYSE variable. If the hypothesized NYSE monitoring is effective, it should partially replace or offset the demand for audit monitoring, thus the demand for higher quality auditing. This should lead to a negative coefficient. (In most cases this coefficient is insignificant, so it is not possible to resolve this point from the research.)

WHY DOES THE ANALYSIS FIND NO EFFECT?

One possibility is that there is no demand for quality differentiation. Another possibility is that there is a demand for differentiated quality, but Palmrose was unable to discover it due to one or more of the following:

1. Inappropriate specification of the independent variable (commented on above). If the Big Eight or industry-specialist splits do not provide a wide enough dispersion of quality to lead to significant coefficients, the author might consider a different calibration of the audit-quality variable. For example, the audit firm name is only one way to differentiate quality — which is a surrogate for the level of assurance that there are no material errors in the financial statements. However, a single firm delivers differential quality with separate signals: "compilation," "review," and "audit." Using this split — alone or in conjunction with the audit firm name — might increase the dispersion on the quality dimension, thus making the regression coefficients more likely to be significant. Obviously, this would require a different sample, because the sampled firms in the study presumably required audits by statute.

2. Inappropriate specification of the independent variables (the factors chosen or their scaling).

3. Incomplete analysis.

4. Inappropriate sample. Since many of the firms in the author's
sample probably require Big Eight firms for reasons unrelated to audit quality, a different sample could be chosen. Palmrose mentions another research project using initial public offerings, which also did not achieve significant results. A possible reason might be that underwriters prefer Big Eight firms (as already noted, for reasons other than audit quality) and encourage new registrants to change to Big Eight auditors if they do not already use them. However, private companies hiring auditors might provide a better sample in that they have a more unrestricted choice among auditors. Obviously, it would be more difficult to obtain information for such companies than for public companies.

THE NEXT STEP?

As already noted, any result of this type of research other than a negative relationship between agency cost and demand for quality-differentiated audit services would have shed little light on agency-theoretic predictions. Since the probability of negative results appears low, the expected payoff of this type of research may be relatively low. Nevertheless, I have been interested in agency-theory work related to auditing since David Ng began investigating the area around 1975. I certainly do not suggest that this empirical work — of relatively recent origin — be strangled in its cradle.

Agency theory still is at a level quite abstracted from the real world. To keep the models tractable, researchers typically make such unrealistic assumptions as single owner, single manager, and single period. The real world is more complex along all these dimensions. Perhaps marginal research effort would more profitably be directed to developing the middle ground of theory before jumping all the way from the mountain top of agency theory down into the trenches of empirical research.

NOTES

1. Peat, Marwick, Mitchell's audit gauge (G) is G = 1.6 x (greater of assets or revenues)²/³ (Elliott 1983).

2. The SEC Practice Section of the AICPA requires that member firms report litigation alleging deficiencies related to SEC audits. That requirement began for suits filed after November 1979. From that date through 30 June 1984 (56 months), 95 suits were reported. Additionally, about 10,000 SEC registrants were audited by section members during that period (the numbers were not reliably determined for the section's first several years, but they were 9,618 as of 30 June 1982 (Public Oversight Board 1983), 10,147 as of 30 June 1983 (Public Oversight Board 1984), and 11,366 as of 30 June 1984 (Public Oversight Board 1984). Thus the average annual
litigation probability for an SEC registrant was approximately .002(95x12/56)/10,000.

REFERENCES


The Demand for Quality-Differentiated Audit Services in an Agency-Cost Setting: An Empirical Investigation: A Reply

First let me say that the discussants' comments should prove very useful for anyone interested in this area of research. The discussants have brought out a number of issues, both conceptual and methodological, that should guide future research. In addition, their comments provide some insight into the problems and trade-offs in doing empirical work of this nature.

I wish to furnish some additional information on just one issue mentioned by the discussants, that is, the sample. Prior to selecting the sample, I was concerned about the potential effect on the analysis, of industry differences in the independent variables. While there are several ways of controlling for any such differences, I decided to conduct the analysis intraindustry. Once making this decision, it was rather difficult to find industries, with adequate numbers of companies having non-Big Eight auditors, for which information was publicly available via proxy statements and annual reports. From a preliminary assessment of data availability, it appeared that office equipment and
retail trade offered reasonable sample sizes. This was not the case for the regulated utilities. These industries were included primarily for the industry specialist/nonspecialist analysis. With hindsight, it seems that a matched-pairs design would have been a more desirable approach to this problem and Dopuch’s comments include some suggestions in this regard.
I congratulate the 1984 Symposium organizers, presenters, and discussants for a very successful and stimulating conference. In my judgment it is the best overall of the six in a series that is now ten years old.

The 1984 conference had two papers related to the audit risk model (Dacey and Ward 1984, and Willingham and Wright 1984); two concerned with substantive statistical sampling (Duke, Bailey, Ko, and Nachtsheim 1984, and Leslie and Aldersley 1984); a paper that explored and summarized the uses of auditor judgment (Gibbins and Emby 1984); and one that was concerned with the economic environment of auditing (Palmrose 1984). The sixth conference was unusual in the variety and quality of the papers, and the importance of the topics covered. Furthermore, as has become a tradition for the symposia, the quality of the discussants' papers was very high.

My job is to synthesize the conference. I plan to do this by reviewing and analyzing parts of the symposia for the past ten years. Specifically, I want to relate the current papers to the comments of Carmichael (1974), Mautz (1974), and Burton and Sampson (1974) from the first conference; Kaplan (1976) from the second; and Groves (1978) from
the third. Also, I want to comment on the current state of auditing research.

HISTORICAL REVIEW

In 1974 the world seemed so young with respect to scholarly auditing research. Statement on Auditing Procedures (SAP) No. 54 (AICPA 1972) was less than two years old, the Elliott and Rogers (1972) paper was barely two years old, and Peat, Marwick, Mitchell’s Research Opportunities in Auditing Program had not yet been created. Two of the 1974 symposium’s presenters and one discussant stressed the need for auditing “research.” I think it is useful to recall the comments made at these early conferences to see what has changed and what has not, and to examine which predictions came true and which did not.

In the first paper of the conference in 1974, Carmichael claimed that the lack of auditing research was due to its “lack of academic respectability,” that “Auditing research has not been fruitful for demonstrating productive scholarship because other members of the academic community have regarded auditing as mechanical, routine, and uninspiring.” Much the same charges could have been made of accounting research, in general, at the time. Empirically-based research methods in accounting research and training in scientific methods in Ph.D. programs were only about ten years old.

The second speaker at the 1974 conference was Mautz, who was Carmichael’s discussant. Mautz attributed the lack of auditing research to (1) a misunderstanding of the nature of auditing, (2) a lack of competence to visualize and perform research in auditing, and (3) restricted availability of research data. These defects suggested to him that joint research by practitioners and professors was needed. (He also suggested that a university-affiliated auditing research center would be beneficial but said that he was too old to head it up. Mautz retired this year after heading up the Paton Accounting Center at the University of Michigan for five years.)

Further, Mautz identified some of the problems facing practicing auditors which he viewed as prime subjects for academic research. He suggested, for example, that “we know very little about how audit conclusions are reached.” Second, that “Recruits come into CPA firms from a great variety of schools with no standardized educational approach.” Finally, he suggested that the introduction of competitive bidding on a national basis tended to increase the pressures upon auditors to “modify their programs in one way or another.” Together these three factors “tend to give us a variety of
solutions where we really need some degree of standardization and uniformity. Thus, one important research question was "do we have a satisfactory degree of standardization and uniformity in audit practice?"

Burton and Sampson also stressed the need for auditing research at the 1974 conference. In particular, they identified three areas in need of inquiry: audit objectives, data-gathering techniques, and human behavior with respect to auditing and reporting. By audit objectives they meant the definition of the economic and legal role of audits, the purpose of the audit, and the costs and benefits of auditing. With respect to data-gathering techniques, they called for research into the amount of evidence needed to support an audit conclusion as well as research into the reliability of audit procedures such as confirmations.

Perhaps most interesting was their call for behavioral research. Rather than taking a narrow view focusing only on auditors' behavior in conducting audits, they also stressed the need for research into the behavior of client employees and client management (that is, anticipatory behavior), and the behavior of users of the auditor's report. With respect to audit reports, they asked, what is the proper interpretation of negative assurance reports, review reports, disclaimers, and qualifications of various types? It also is interesting that they called for research into several problems that have since been "resolved" by the Auditing Standards Board by issuing Statements on Auditing Standards (SASs). Specifically, they asserted the need for research into problems of related parties, reliance on other auditors, and auditor participation in reviews of interim financial statements.

At the second symposium, Kaplan defined (scientific) research and contrasted it with development of products. He illustrated his points with research in statistical sampling and the development of computer audit software strategies and noted the comparative advantage of professors in conducting research. Like Mautz, Kaplan called for joint efforts by professors and practitioners. He also discussed the lack of knowledge of how auditors behave and called for what might be described as a "case study" approach to developing a knowledge base to assist in research in auditing.

Finally, in 1978, Groves echoed Mautz's view, focusing on the inadequate background of professors in regard to auditing research, and reinforcing the perception that the data necessary to perform major audit research were not and would not be available to academia. Like Mautz and Kaplan, he suggested that joint work might offer a solution. Groves also was concerned with what he believed to be the
detrimental effects of “excessive” competition. He said, “The line between a truly efficient audit and one that is economized beyond the point of adequate scope is a fine one. Cutting corners . . . is an almost unavoidable reaction to fee pressures.” Like Mautz, Burton, and Sampson, then, he was concerned with the proper amount of auditing.

Thus, we see four themes: (1) the need for research in auditing (however research is defined); (2) the perceived lack of respectability of auditing research on campus; (3) the alleged lack of qualifications or data for professors; (4) the concern with the average quality, or perhaps the standard quality of auditing, and the effects of competition.

The early commentators did not discuss generally accepted auditing standards (GAAS) per se, yet standards seemed to be a driving force. I believe that it would surprise members of other learned professions to know the extent to which professional standard setting for auditors leads both practice and research in the field. For example, it is my view that the publication of SAP No. 54 in 1972 exceeded the then prevailing practices of almost all practitioners. The issuance of this professional standard, more than any other single event, was responsible for generating academic respectability for studies on auditing procedures. This was achieved by providing a rich conceptual basis for audit planning and audit conclusions. Specifically, it delineated sources of audit reliance, types of evidence, how they are related, and how their results may be combined. The concepts of internal control and their relationship to substantive tests were explained, as was the combination of reliance on subjective evaluations and judgments with more objective statistical samples and other sources in a quasi-Bayesian framework.

SAP No. 54 was not just a theoretical treatise, but also was the standard of the profession. It led to innovation in practice and was quickly incorporated into textbooks, thereby becoming a part of the common training of the next generation of professors and auditors. It was, indeed, a major development affecting practice and academic respectability for auditing, and ultimately it helped to attract young professors to auditing as an area for research.

It is interesting that the concern of early presenters with standardization, the floor of audit quality, and “excessive” competition preceded the most controversial auditing standards (specifically, SAS No. 39 and SAS No. 47). Yet little scholarly work has been done on the effects of changes in competition in auditing that may be related to relaxation of ethical prohibitions and their relationship to the demand
for professional standards. Also, there has been little work on the effects of the standards themselves.

CLASSIFICATION OF PAPERS

Against this historical background, I wish to classify the papers presented at the six symposia held over the last ten years. In the Chart we see the domains of accounting and auditing, and their intersection. Together they determine the value of a system of audited financial reporting.

The left-hand side, or the accounting domain, is comprised of how accounting is done — that is, the methods of accounting. Accounting standard-setting bodies have the responsibility to determine acceptable and unacceptable methods. Consequently, auditing has little to do with this side. The middle (lens) area is the intersection of accounting and auditing. It is comprised of the accounting and auditing concepts of materiality. Finally, the right-hand side is comprised of auditing procedures and auditor reporting and has little to do with accounting.

Chart. Classification of Symposia Topics

3 General research reviews
4 Environment of auditing
or materiality per se. The auditing side is divided by a horizontal line that separates audit-procedure from audit-reporting issues. With respect to the materiality region, the horizontal line separates the ex ante fact-finding or "discovery" materiality from ex post, or after-the-fact "disclosure" materiality. It should be noted that as we move from left to right and from top to bottom, each subsequent component is conditional on the results or specifications to the left. For example, materiality may be conditional on the accounting methods used while the design of auditing procedures is conditional on the level of materiality specified as well as the accounting methods used. Also, of course, whatever might be reported is conditional on the precision of the investigation based on discovery materiality.

There have been 39 papers presented (not counting the dinner speeches) at the six audit symposia (see the Table for classification of all papers presented from 1974–84). Seven of these do not fit within the classification scheme of the Chart. Three were the early papers calling for research in auditing and four were concerned with auditing at a more aggregated environmental level. For example, the Palmrose (1984) paper encompasses both the accounting and auditing aspects of audited financial reporting.

Of the remaining 32 papers, only one can be classified as a "discovery materiality" paper (Leslie 1976). It was probably the most innovative and important single paper of the series, yet it is among the most ignored. Two papers can be classified as "disclosure materiality" and perhaps three as audit-reporting papers. The remaining 26 are concerned with audit procedures! Only four were joint products of practitioners and professors and only two of the four involved data analysis. One of the data-based joint papers was presented by Willingham and Wright at the 1984 symposium.

About five-sixths of the papers have been devoted to the relatively invisible and constrained subject of audit procedures applied to obtain sufficient, competent, evidential matter. These can be subdivided as shown in the Table. Almost 60 percent relate to quantitative auditing topics with auditor behavior the only other major category. The analytical review papers are particularly noteworthy, since they comprise a fairly large proportion of the major papers published in the field to date. But, there has been almost nothing presented on audit reports and their uses, nothing about how much auditing is enough or where GAAS should be set, and no discussion of what legal liability rules should be. Similarly, there have been no behavioral studies of how client employees or managers alter behavior in anticipation of audits (or how such behavior changes as audit scopes are altered
<table>
<thead>
<tr>
<th>Sampling (S)</th>
<th>Analytical Review (A)</th>
<th>Audit-risk Model (R)</th>
<th>Auditor's Behavior (B)</th>
<th>Other (0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discovery Materiality (MP)</td>
<td>Disclosure Materiality (MR)</td>
<td>Audit Reporting (AR)</td>
<td>General Discussion of Research (G)</td>
<td>Environment of Audited Financial Reporting (E)</td>
</tr>
</tbody>
</table>
under competitive pressures), how management considers auditing in reducing agency costs, nor how audit reports affect users via the stock market. Finally, there has been only one paper that even touches on the extension of auditor attestation services beyond historical audits.

Thus, while progress has been made in the areas covered, we seem to have been unnecessarily narrow in our view of what constitutes appropriate auditing research. This leaves us with considerable opportunity for future work that I would like to sketch out briefly.

CURRENT OPPORTUNITIES IN AUDITING RESEARCH

Under the assumption that the symposium papers are reasonably representative of the auditing research literature, it seems clear from the Chart that the most underdeveloped areas are those located on the main diagonal. It is not an accident that they are almost empty — they are difficult areas to research, yet there is much opportunity here. Obvious candidates are the areas of materiality and audit risk. For example, what does SAS No. 47 mean in terms of ex ante audit planning and ex post liability? How low is "acceptably low" audit risk? How much is "material" and how should it be allocated among accounts (in both an accounting and auditing sense)? What does SAS No. 47 mean in an efficient capital market, and what implications does it have for agency costs, monitoring, and auditors' legal liability?

Similarly, research which would fit into the reporting cells would be fruitful. For example, what is the information content of the audit report, review report, forecast review, or compilation report? (Note that there are, of course, nontrivial problems of separating the effects of the auditor's report from the numbers themselves.) There also is much opportunity for research on the main cell of audit procedures. It is not likely, however, that there will be further division of the focus of behavioral studies into microscopic studies of auditor decision-making (the Gibbins and Emby (1984) paper suggests many new areas for broader studies). Similarly, I believe that there is relatively little opportunity remaining for development of ever better dollar-unit sampling (DUS) evaluation techniques (there are still some out there, but it is not a high expected-pay-off area).

Substantial scholarly contributions may come from operationalizing more integrated approaches to problems of economical accumulation of evidence for the audit as a whole. A large part of the early research in audit sampling involved exploiting the structure of accounting data on an account or transaction stream basis. Today, we can exploit
our knowledge of a firm's industry and operation and the structure imposed by an accounting and control system in assessing audit risk over the statements as a whole. The microcomputer can assist audit field workers by (1) keeping track of interrelated data and interrelated audit conclusions, (2) providing decision aids to assist with complex judgment tasks, and (3) applying structural equations and Bayes' rule to integrate audit evidence from various phases of the audit. A microcomputer-based system also could be used to collect data for a firm-wide data base with which to set policy guidelines. (The Willingham and Wright (1984) paper is illustrative of some possibilities for the latter.) The microcomputer makes these advances possible and competition will make them necessary. We professors have a great opportunity to do the research necessary to proceed with the development of improved integrated methods.

Finally, as professors interested in auditing, we need not limit ourselves to auditing procedures. The Palmrose (1984) paper is an illustration of research into the economics and environment of auditing. An understanding of the institutions and nature of auditing is necessary for making progress in these areas. We should take advantage of the opportunity to use our institutional knowledge of auditing to resolve these problems even if it means working with other scholars who have skills in numerical analysis, economics, or finance.

In closing, we have come a long way in ten years and the contributions of the Illinois symposia have been considerable. They have provided a forum for the exchange of ideas about auditing through the formal program as well as through the informal discussions among leading practitioners and professors in a congenial atmosphere. We in the audit area are grateful to all of those who contributed their efforts and support to make these six symposia successful.

REFERENCES*


W. S. Albrecht and J. C. McKeown, "Toward an Extended Use of Statistical Analytical Reviews in the Audit." Symposium on Auditing Research II

* All dates shown in the reference list for symposium papers are the dates of the actual conference meeting rather than the date of the published proceedings.
American Institute of Certified Public Accountants, Statement on Audit Procedure (SAP) No. 54, "The Auditor's Study and Evaluation of Internal Control" (November 1972).


———, Statement on Auditing Standards (SAS) No. 47, "Audit Risk and Materiality in Conducting an Audit" (December 1983).


D. A. Johnson, “To Blow the Whistle or Not: An Employee’s Dilemma with Internal Control Implications,” *Symposium on Auditing Research* III (Urbana-Champaign: University of Illinois, Department of Accountancy, 1978), pp. 239–56.


———, “The Roles for Research and Development in Auditing,” *Symposium on Auditing Research* II (Urbana-Champaign: University of Illinois, Department of Accountancy, 1976), pp. 3–11.


——— and S. J. Aldersley, “A Model for Multi-Location Audits,” *Sym-*


(Urbana-Champaign: University of Illinois, Department of Accountancy, 1974), pp. 23–59.

