*Accounting Horizons*
Vol. 28, No. 3
2014
pp. 695–712

American Accounting Association
DOI: 10.2308/acch-50818

**COMMENTARY**

# Some Methodological Deficiencies in Empirical Research Articles in Accounting

**Thomas R. Dyckman and Stephen A. Zeff**

**SYNOPSIS:** This paper uses a sample of the regression and behavioral papers published in *The Accounting Review* and the *Journal of Accounting Research* from September 2012 through May 2013. We argue first that the current research results reported in empirical regression papers fail adequately to justify the time period adopted for the study. Second, we maintain that the statistical analyses used in these papers as well as in the behavioral papers have produced flawed results. We further maintain that their tests of statistical significance are not appropriate and, more importantly, that these studies do not—and cannot—properly address the economic significance of the work. In other words, significance tests are not tests of the economic meaningfulness of the results. We suggest ways to avoid some but not all of these problems. We also argue that replication studies, which have been essentially abandoned by accounting researchers, can contribute to our search for truth, but few will be forthcoming unless the academic reward system is modified.

**Keywords:** research methodology; statistical analysis.

Methodological discussion, like calisthenics and spinach, is good for us.
—Paul A. Samuelson

*Thomas R. Dyckman is a Professor Emeritus at Cornell University and an Adjunct Professor at Florida Gulf Coast University, and Stephen A. Zeff is a Professor at Rice University.*

We are particularly grateful for the comments received on earlier drafts from two anonymous reviewers. We also thank Kees Camfferman, Hans Christensen, Steve Crawford, Paul Griffin, Andrew Jackson, David Johnstone, Mike Minnis, Mark Nelson, Ken Peasnell, Terry Shevlin, Shiva Sivaramakrishnan, Richard Swartz, and Bob Swieringa for their comments.

Italics used in quotations appeared in the original. We have added the contents within brackets in quotations except where noted.

Supplemental material can be accessed by clicking the link in Appendix C.

*Submitted: September 2013*
*Accepted: May 2014*
*Published Online: May 2014*
Corresponding author: Stephen A. Zeff
Email: sazeff@rice.edu

## PREAMBLE

In this paper, we undertake a critical review of the research methodology used in a sample of 66 papers recently published in two major accounting journals. We draw attention to deficiencies in the authors' application of methodology to empirical research, and we propose remedies which accounting researchers may wish to take into account in their future work.

After an introduction in which the sample of articles selected for analysis is described, the paper continues by addressing several methodological deficiencies we believe are common to current empirical accounting research. We suggest possible ways of avoiding, mitigating, or moderating the difficulties identified.

The development of this paper unfolds as follows. In the introduction, we describe the set of articles used in our analysis. We then turn our attention to concerns associated with the authors' selection of the sample period in our sample of articles. The next section is devoted to our concerns over how the data in the articles were obtained and analyzed. Our discussion is primarily focused on authors' adoption of their statistical methods. This section is divided into two subsections. The first subsection covers those papers employing a regression framework. The second subsection considers the behavioral research studies included in the journals we examined. The next section includes additional suggestions of approaches that can ameliorate some of the problems currently endemic to regression studies. We then present an analysis of how the results are summarized in the "Conclusions" sections of the articles, together with how we believe they *should* be summarized. We conclude by addressing the implications for authors, their readers, editors, reviewers, and educators in general. We alert the reader that our own views on how to improve the current state of affairs arise throughout the paper.

## INTRODUCTION: OUR DATA SET

Authors, editors, and reviewers of manuscripts submitted for publication in academic journals are collectively responsible to their readers for the quality of the published work. We maintain that authors in particular, but also editors and reviewers, have an ethical duty to fulfill these responsibilities. We do not question intent, but we do question how this task is almost universally discharged. In particular, we focus attention on studies that subject time-series or cross-section data to regression analysis and also on a set of behavioral studies. These studies have become increasingly popular in recent years, and they now constitute a substantial percentage of the recent articles published in the leading journals.

We have reviewed 66 articles that constitute 90 percent of the total articles published in two prominent journals from September 2012 through May 2013. The remaining 10 percent are analytical studies, which we do not address. Of the 66 papers, 11 are behavioral, which we cover separately. The remaining 55 papers are empirical studies involving regression analysis. We cite examples of the methodological weaknesses we have found. The articles, listed by the journal issue in Appendix A (empirical studies) and Appendix B (behavioral studies), appeared in the September and November 2012 and the January and March 2013 issues of *The Accounting Review*, and in the September and December 2012 and the March and May 2013 issues of the *Journal of Accounting Research*. The appendices appear at the end of the article and, together with supporting data for a number of our empirical counts of articles shown below, are available in the Supplementary Data File from the American Accounting Association (hereafter, SDF). The SDF can be accessed by clicking the link in Appendix C. These groups of articles from two of the leading accounting research journals were selected as being representative of recently published research. The data we cite on these articles have been obtained by our direct examination, not by using an electronic or other search engine.

We are concerned with both the integrity of the sample-period data as well as with the quality and the reporting of the data analysis. We will suggest approaches to improve both the relevance of the research and the identification of constraints to improving the publication process.

## SAMPLE PERIOD SELECTION

An indispensable question to be addressed in any paper, once the authors have specified the sample period, is: what assurance can readers have that the sample period is reasonably stable itself and in relation to the periods in close proximity. This should be the case so that any test is not distorted by factors specific to, or even close to, the period selected but not germane to the analysis. The research findings are inevitably the product of the research design, and a key element in the design is the selection of a sample period that is free from external bias and other confounding factors. Why did the collection of research data begin in year x and end in year y, as opposed to some other sample period? Sometimes the starting or ending point may appear apparent to a reader, as is the case in event studies. But even here the period selected must be carefully chosen so that the event of interest is not affected by extraneous events. A recent editorial in *Psychological Science* spells out the requirements that must be acknowledged by the author of a paper submitted for publication and which must include the "stopping rule" used for data collection (Eich 2013). We interpret this to mean both the starting and stopping rule in order to justify both the time period selected to start data collection as well as to end data collection.

Consider the requirement to report under IFRS. Should the study necessarily begin in the year in which the relevant data *first* became available? And in what year should the study end, other than as might be decided arbitrarily by the researcher to end just as the study is being conducted? This data availability criterion may be described as a "constraint." Yet the first year in which the data became available does not *necessarily* qualify as the appropriate first year of the sample period, and the last year in which the data are still available does not *necessarily* qualify as the appropriate final year of the sample period. Suppose, for example, the first one or two years, or the last one or two years, of the period in which the data were available were of dubious stability because of, say, the passage of major legislation, the approval of important regulations, the issuance of major amendments to relevant accounting standards, the withdrawal of relevant accounting standards, the occurrence of a macro-economic shock, or an increase in managerial actions taken in anticipation of eventual adoption.[1]

Of the 55 regression papers we reviewed, only ten provided a justification of any kind for the selection of the beginning year in the sample period[2] (#s 2, 3, 7, 17, 34, 40, 45, 50, 51, 53), and in every instance the selection was defended in terms of data availability alone.[3] Only three papers (#s 15, 34, 35) justified the selection of the ending year, again for reasons of data availability. But this constraint says nothing about whether the sample period selected is the one that captures just the impact of the issue being studied, without being sullied by micro or macro shocks. In the 55 papers examined, we did not find a single instance in which the authors undertook to assure readers, through suitable argumentation, that their sample period was indeed appropriate for analyzing the research question. Without such a defense of the sample period, how can readers have confidence that the sample period is anything more than an arbitrary selection, an aberration? The lack of *any* defense of the propriety of the sample period in papers—the authors of almost 80 percent of the

[1] This problem is endemic to event studies where the event itself, or shock, is unlikely to be independent of the prior history of the dependent variable. The researcher must be able to argue persuasively that both the size and timing of the event are exogenous. See Hamermesh (2000, 370) on this point and related topics.

[2] The numbers correspond with the numbered articles in Appendix A.

[3] An interesting case is provided by article #3, where the authors did explain that they postponed the beginning of the sample period by two years from the year in which the data first became available, because the early data were insufficiently refined for the authors' research.

papers we examined said absolutely nothing about why the particular sample period was chosen—represents a blind spot in the accounting literature,[4] and it might impair the ability to generalize the research findings.

In other literatures, a question as to whether the sample period may be unstable and therefore not a sound basis on which to conduct the research may be tested by conducting a replication. By replication, we mean redoing the identical study in the same way but for another sample period, or periods.[5] But in the accounting literature replications are not rewarded by either authors' colleagues or their academic institutions. Status and promotion go to the researcher who is first to publish, particularly if the methodology is novel and impressive. The reward structure therefore encourages researchers to conclude their work even though the rush to submission may introduce errors. This process, if anything, enhances the need for, and importance of, replications. Recently, Philip Brown (2013, 856), a distinguished accounting academic, made a plea that editors and reviewers should refrain from discouraging replications.

Without replications, researchers are likely to accept, simplistically and without question, the results of the previous studies on which the current study depends. A relatively recent paper by Bamber, Christensen, and Gaver (2000) provides an excellent case of undertaking a replication study, which is their analysis of Beaver's seminal 1968 paper, "The Information Content of Annual Earnings Announcements." Even more important, they describe their own work in replicating Beaver's approach on a set of firms using the same approach. They appropriately adopt the critical stance that "The review process in accounting suffers from a reluctance to publish replications and a prejudice against the null [hypothesis]" (Bamber et al. 2000, 123) and also suggest that the initial publication of results can act as a deterrent to future confirmation studies.[6] Indeed, the articles in our sample almost uniformly avoid mention of any prior *confirmation* of the results of the previous studies cited.[7] We also find it surprising that this important paper by Bamber et al. (2000) was not published in the journal that published Beaver's paper.

The value of well-done replications should be appreciated not only by authors but also by editors and reviewers. A positive replication would suggest that the result is well-grounded and can be relied upon or quoted. Furthermore, future efforts might perhaps be better directed toward making extensions or to addressing new topics.[8] A failed replication, on the other hand, would warn

[4] We have been present at workshops where the authors apparently viewed their selection of the sample period as so inconsequential to their research design that it was revealed, not in the text of the paper, but in a footnote or in the small print in an exhibit in the back of the paper. Also in workshops, typically involving junior faculty or doctoral candidates presenting their road paper, we hear the standard defense to criticisms of methodology, namely, that "it is in the literature." This is no justification whatsoever, for, as we point out in this article, there is a great deal of questionable use of methodology in the literature. Novice researchers should be disabused from unthinkingly relying on previous literature for guidance on sound methodology.

[5] This approach will not be possible in an event study. We also acknowledge that most of the papers we examined included robustness tests that provide improved precision, additional perspectives, and the investigation of potential alternative explanations.

[6] The blossoming of empirical research in accounting can be traced back to 1968. The seminal papers by Ball and Brown (1968) and Beaver (1968) can be likened to the appearance of two Black Swans (Taleb 2007) in their impact on accounting research. Unfortunately, the bias against replications prevented realization of the full advantage from the generally accepted research process in accounting that we would describe, borrowing from Taleb, as "antifragile" (Taleb 2012).

[7] Although few in number, critiques of the seminal papers by Beaver (1968) and Ball and Brown (1968) were available as early as Chambers (1974), Dyckman, Downes, and Magee (1975), and, later, empirically in Oppong (1980) and Atiase (1985).

[8] A case of inappropriate reliance is suggested by the work of Merino, Koch, and MacRitchie (1987), in which the authors cast serious doubt on the validity of claims made by Chow (1983) in his AAA manuscript award-winning paper, which was then cited as confirming evidence by Watts and Zimmerman (1986, 173) in their book, *Positive Accounting Theory*. While Merino et al. (1987) did not constitute a replication of the Chow paper, the lesson is the same. The additional fact that the Merino et al. paper was published as a comment rather than as a full article is an indication of the appreciation for replications, at least in the 1980s.

the reader not to accept, quote, or use the results as a basis for extension until the matter is clarified. Unfortunately, we speculate that the number of replicated studies relied on in our set of papers is close to zero.

We suggest that journals' editorial policy should require the general availability of all the data for any published study, as well as a description of the research program, in order to allow analysis of the work and the potential for replication of the study. Readers and future authors will expect that inaccurate data, or data that do not support the null hypothesis, will not be purposely omitted, and that the data have not been fabricated to fit a preconceived hypothesis. Unfortunately such instances are not unknown even in the natural sciences. Furthermore, researchers, including accounting researchers, are not known for their willingness to make their data available to requests from their colleagues.[9]

Unfortunately, there is little or no "market" for replications in our literature, resulting in its being largely devoid of replication studies that would lend credence to prior work.[10] Universally in the sciences, replication is regarded as a hallmark of the scientific method. A deputy editor of *Science*, the weekly journal of the American Association for the Advancement of Science, together with three co-authors, has recently called replication "the scientific gold standard." They added, "The importance of replication and reproducibility for scientists is unquestioned" (Jasny, Chin, Chong, and Vignieri 2011), yet replication is eschewed in the accounting literature.[11] Without replication, it becomes *essential* that the author of the original article adequately defend the propriety of the sample period.[12] One approach would be to adopt a hold-out sample strategy as a means of ascertaining integrity, for example, by using alternate or a random set, or sets, of years to lend credibility to the stability assumption. An alternative strategy would be to consider examining one or more prior or subsequent periods so that the implicit assumption of process stability achieves additional credibility.

Perfect stability of an economic process is not possible, but a reasonable level of stability is essential and needs to be justified. Researchers commonly use a zero/one variable in their regression

[9] Documentation of these two statements involving even some of the more famous researchers in the natural sciences can be found in Broad and Wade (1982), Hartmann and Moers (1999, 2003), and Bailey, Hasselback, and Karcher (2001).

[10] An indication of the importance of replications is evident in a study by Begley and Ellis (2012), which examined 53 landmark studies in the fields of hematology and oncology only to find that just six (11%) could be successfully replicated. In addition, two articles recently appeared in *The Wall Street Journal* (Naik and Martin 2014; Martin and Naik 2014) which reported that a new safer, easier, and more ethical approach to creating stem cells, which was discussed in two papers in *Nature* (Obokata, Wakayama et al. 2014; Obokata, Sasai et al. 2014), has not been validated by at least ten attempts at replication. Indeed, one co-author has requested that the lead author retract both papers (Martin and Naik 2014). Recently, the Many Labs Replication Project re-ran thirteen psychological experiments concerning widely accepted theories, of which ten were validated (reported in Combating bad science: Metaphysicians 2014).

[11] The October 19, 2013 issue of *The Economist* (Unreliable research: Trouble at the lab) contains a lengthy critique of scientific research. On page 26, it states as follows: "The idea that the same experiments always get the same results, no matter who performs them, is one of the cornerstones of science's claim to objective truth." But change is underway. On page 30, *The Economist* reports that Brian Nosek, a psychology professor at the University of Virginia, has announced a new Center for Open Science to "make replications respectable," and it notes that one of the flagship publications of the Association for Psychological Science, *Perspectives on Psychological Science,* will soon have a section devoted to replications. The March 15, 2014 issue of *The Economist* (Combating bad science: Metaphysicians) describes a new laboratory, METRICS, which will address the issue of reproducibility in medicine, statistics, and epidemiology as well as publication bias, which can lead to over-emphasis on results reported as statistically significant. One possible explanation for this over-reliance is provided by Kuhn (1962, 35) and quoted by Bamber et al. (2000, 111): "And the project whose outcome does not fall in the narrower range [of anticipated outcomes and thus assimilable results] is usually just a research failure, one which reflects not on nature but on the scientist" (brackets in the original).

[12] We believe that the paucity of replication studies is also an important contributor to the research issues raised later in this paper. However, we note that three partial replications do appear in the studies examined (#s 8, 46, 49).

to account for one or more stability issues within the regime. If there is truly a change, but it is improperly identified, the related instability will remain, resulting in an averaging of the effect on the intercept and the slope coefficient. Indicator variables were used in 52 of the 55 regression studies we examined (#s 22, 49, and 53 were the exceptions). When a change in stability can be correctly identified, its effect can be incorporated in the analysis, thus retaining reasonable stability. Because reasonable stability is essential, the choice of the variable and the assignment of the indicator values should be adequately justified. Merely defining when the variable takes on the value of one or zero, as is generally the case in the current studies, is not sufficient.

Another way to address the problem of stability is to subdivide the period and see if the separate regression analyses yield similar regression coefficients. Unfortunately, shorter time periods could lead to missing data, and event studies often do not have sufficient post-event data to test, because of the researcher's desire for reportable results. But even when adequate data are available, the internal stability problem, while admittedly not easily resolved, is seldom acknowledged in these papers. In event studies designed to examine before/after effects, the hypotheses should indicate those aspects of the post period that should show change and those that should remain essentially unchanged. We identified 17 studies, listed in the SDF, where we believe events of this type would be likely (see, for example, #s 4, 38, 43).

The point of this discussion is that authors must resist the temptation to treat the selection of the sample period casually, as if it were of little or no consequence to their findings. We recommend the use of a hold-out sample strategy as one possible means of achieving integrity.

## ARTICLES: DATA SELECTION AND ANALYSIS

The advent of very large data sources, including Compustat, CRSP, and EDGAR, plus computerized statistical software packages, has led to an explosion of statistically oriented papers in accounting and elsewhere. Unfortunately, the ability to run classical null hypothesis statistical tests (NHSTs) has often been accompanied by a failure to recognize that consideration of the economic significance of the investigation is its primary purpose. The research program should not be driven by the availability of a specific data set or adaptable computer program but instead by the research question(s). We will have more to say about the sampling process itself after examining aspects of the data analysis.

### Data Analysis: Randomness and Tests of Significance in Regression Studies

The regression papers in our study typically examine an existing data set that relates to the researcher's posed research. The selection of the topic, which should precede the data-gathering process, is described in the paper's introduction.[13] An early activity is to assure that the data are rigorously checked for correctness. In addition, the author should document what has been done in this regard. Special attention needs to be given to the theory behind the research selection, not only of the topic but also because it will have an impact on the variables that are to be used and their measurement. In particular, the independent variables should reflect the theory and not be selected based on prior tests of their significance for selection, as is not uncommon. This work needs to be done before the statistical analysis is performed. The actual process to this point is critically important to the reader in evaluating the value of the research, but, unfortunately, is seldom if ever

[13] If the complete research program were required to be submitted in advance, many of the unknown problems in reviewing and ascertaining methodologically rigorous performance could be avoided in the review process, while allowing the reader to sleep more soundly. The editorial process now required by *Psychological Science*, mentioned above and again later, requires several disclosures, including a report on all failed as well as all successful manipulations that relate to the targeted research (Eich 2013).

described in any detail in these studies. The reader is apparently expected to assume that everything was done properly. The process, along with the sample, should, at least, be available to the reader from some accessible source for review and analysis, and for potential replication. If all is as it should be, we have our data set.

The time-series and cross-section papers we examined all use statistical regression to analyze the data. The interpretation of the results relies uniformly on making statements about the statistical significance of the calculated slope coefficients (b values). The mathematics that underlies the calculations is the probability calculus, which requires that the process that generated the data be truly stochastic, which in turn requires that the data be the result of a random (sampling) process. This fact alone has important implications for the research. But is our sample a random sample? Perhaps not, but it is all that we have been allowed by Nature to examine. Can we then proceed as if the sample is random? Will the results be defensible?[14]

A randomness problem arises because accounting deals with observed economic and social data, not data from a randomized experiment, as is the norm in the physical sciences. The data values should not be viewed simply as deviations from some imagined unique regression surface subject to random deviations, as is appropriate in the natural sciences. The observed data for each variable are actual economic or social outcomes. The individual data points in these studies are unlikely to reflect an identical economic process. The reporting of operating income across firms, for example, will differ depending on the firm-specific calculation process. In fact, the available data are actual observations of economic or social events, albeit perhaps inaccurately recorded or recorded under different circumstances. Initially, we need to admit that the data samples we use in accounting and economics in general are not random samples, defined as a sample drawn from a well-defined population by a process that assures that each member of the population has an identical probability of being selected in the sample. Random samples are simply not possible in our discipline. This means that the sample is more appropriately viewed as a population.[15]

Concurring with Johnstone's work (1988, 1997), we believe that a level of randomness can be attained which, though not ideal, should be sufficient to allow the researcher to proceed, albeit with caution. Proper experimental design depends on stratification from which a "random" sample within each stratum can be achieved under specified constant conditions. However, we are unable to document any *explicit* attention to this aspect of experiential design in the papers we reviewed.

Our sample's authors apply the techniques of regression analysis to their data, and they report NHSTs. Unfortunately, the sample data do not constitute a random sample, as required by the probability calculus. If the results are to be reported, the author(s) need to evaluate them with special attention. Johnstone (1988, 324, relying on Kyburg [1976] and the writings of R. A. Fisher), argues that "a test of significance is legitimate, irrespective of whether the sample was drawn at random, if the reference set includes no recognizable (given current knowledge) relevant subset containing that sample."[16] If researchers elect to follow the classical Fisherian approach, that fact should be revealed to the reader. The extra care required under this dispensation demands that the researcher argue persuasively that the reference set restriction noted in the Johnstone quotation has been met. None of the papers we examined do so.

---

[14] A recent paper by Balsam, Irani, and Yin (2012, 411) recognizes at least that, "To the extent the firms not in our sample are systematically different from our sample firms, the reported Ordinary Least Square [sic] regression results would be biased because the sample is not randomly selected."

[15] We admit to, but deplore, the possibility that the availability of a pristine data set can drive the research agenda.

[16] The philosophy here is basically a Bayesian one, hence subjective in nature, for which the sample needs only to appear random, subject to current knowledge. We note that Ziliak and McCloskey (2004) point out that the Fisherian position articulated by Johnstone was not shared by all statisticians. Begley and Ellis (2012) suggest that this may no longer be the case.

An alternative approach to NHSTs recommended by Cumming (2013) suggests the use of confidence intervals rather than the classical probabilities: 10 percent or 5 percent or 1 percent, as is routine in over 80 percent of the papers in our sample. To be sure, one could also simply report the actual probability. The confidence interval, however, retains a clear element of the uncertainty involved, which any single probability number lacks. An interval also suggests the better interpretation that should be given, which is that the interval reflects what one should expect the frequency to be in the future. Unfortunately, neither a NHST nor a confidence interval considers any other information relevant to the analysis or to any decision problem, such as costs. In particular, the reader is not supplied with information as to the likelihood of the most believable alternative hypotheses that have been summarily rejected. While the most believable alternative hypothesis is not always going to be obvious, Johnson's (2013) results show that, if we can assume that one-half of the possible alternative hypotheses should yield a positive result, then this suggests "between 17% and 25% of the marginally significant results are false." The percentages increase if researchers are reasonably good at selecting the null. To allow for this real constraint, Johnson suggests that researchers using the traditional probabilities of 5 percent and 1 percent should instead insist on levels of 0.5 percent and 0.1 percent as a minimum to avoid false positives.[17]

But the fact that our sample is not a random one should not be viewed as an invitation to incorporate other techniques that further distort randomness. For example, it is not uncommon for authors to merge data from different sources (see, for example, #s 1, 51, 55). But these different sources are unlikely to reflect the same underlying process, or they may exclude data for one or more variables or merge data from several sources. We list 30 such cases in the SDF (see, for example, #s 23, 37, 52). The outcome is that the resulting data are still further removed from the conceptual random sample.

A second commonly accepted practice is to exclude certain data considered to be susceptible to error. Winsorizing (defined as revaluing all observations that exceed the 99th percentile or fall below the 1st percentile to equal the 99th and 1st percentiles, respectively) is commonly employed to prevent these observations from having undue influence on the results. We identify 18 such cases in the SDF (see, for example, #s 7, 38, 51) but believe there are additional cases where the authors do not divulge that they have winsorized. The desire to exclude the "outliers" arises due to the researcher's concern that the information content of these data points, if measured with error, will have a disproportionate effect because the regression technique squares their impact. But, assuming proper researcher care, these data points are also actual observations, and excluding them is not appropriate under random sampling techniques. Indeed, these data points are likely to be the more interesting to analyze and possibly represent the intrusion of a changed or different process. If the implicit squaring of the residual is an issue, the researcher can consider instead minimizing an alternative weighting scheme such as using absolute deviations in robustness tests.[18] Another possibility is to run the regression, again with the points included, as one author did (see #28). A recent editorial in *Psychological Science* (Eich 2013) recognizes this *ad hoc* procedure by requiring the author(s) to provide the number of, and reason for, excluding each and every observation at the time the paper is submitted.

[17] We note that many of the regression papers in our sample report either standard errors or "t" statistics that permit the reader to compute confidence intervals. This is true of approximately one-half of the behavioral papers.

[18] See the working paper by Leone, Minutti-Meza, and Wasley (2014), which describes the problems introduced by truncating generally, and winsorizing in particular, even when done properly on both the dependent and independent variables, and a discussion of alternative techniques devised to improve if not remove the resulting bias. Chambers, early on, while he does not mention winsorizing directly, writes of being uncomfortable because "the very act of choosing between parts [of the data set] seems often to mean rejection of some parts as irrelevant" (Chambers 1973, 169).

We suspect that many of the authors in our sample were simply not aware of the need to consider these matters, or assumed—and would no doubt submit—that their samples at least approximately satisfy the statistical requirements noted by Johnstone in the previous quotation. Given this situation, it is important to ask what we could learn from the regression results in these and similar papers. For example, it is possible that the results reported as significant may indicate that the independent variables involved, once benefiting from additional analysis, may be shown to be relevant indicators of effect. Studies involving well-documented data sources such as Compustat, CRSP, and EDGAR provide potential examples of sources for readily available samples (see, for example, #s 4, 15, 55). The studies we examined, however, do not attempt to provide additional justification for their approach. Furthermore, the reported significance levels should be accompanied by a statement indicating that the results are not based on a random sampling process. This statement could be accompanied or replaced by words such as, "the reported significance levels might approximate those that would have been attained if a truly random sampling process had been possible."[19] The authors owe the reader a clear statement concerning the basis and limitations of their work.[20]

But there is a more important issue to address. A statistically significant result is not necessarily an important result. Without establishing the *economic importance* of the result, which requires additional work on the part of the researcher, the mathematics reported to date is worthless. Examples of variables used in the studies we examined that could reflect the importance of the research findings include the change in CEO compensation and the cost of equity. The importance of establishing the economic significance of any result is strikingly illustrated by an example from Ziliak and McCloskey (2004, 530): "a statistically *in*significant coefficient in a financial model, for example, may nonetheless give its discoverer an edge in making a fortune; and a statistically *significant* coefficient in the same model may be offset in its exploitation by transactions costs."

Only a quarter of the papers we examined recognize the critical importance of estimating the economic significance, even incorrectly, of their research (see, for example, #s 1, 33, 39, and the SDF). And of the papers that do attempt to do so, only two appear to consider the economic significance sufficiently important to include an estimate thereof in their Conclusions sections.[21] We believe that the "Conclusions" section needs to be more complete and feature a change in focus toward the economic significance of the findings. We urge reviewers and editors to consider the reporting of the economic importance of the results as critical to final publication acceptance. Whatever the reasons for current omissions, we submit that the current cursory attention, where present, represents a major disservice to the interested reader and therefore to the profession at large.

The accounting academy, however, appears to be more concerned with statistical significance than with the economic importance of the results reported. As suggested above, a statistically significant result is meaningless in and of itself. The additional and essential step, then, is to be able to establish the economic importance of the finding. Importance dominates fit and statistical significance.

Often the research neglects the Type II error. The author may not have spent sufficient time to assure that the results are sufficiently powerful against alternative reasonable hypotheses. In other

[19] This statement paraphrases the suggestion by Winkler (2009, 190).

[20] The fact that the sample is not a random sample does not prevent the researchers from calculating descriptive measures of their basic data. Indeed, nearly half of the papers in our sample do so. The basic data are social science data and therefore are unlikely to be distributed in accordance with a normal distribution. Hence, these values are not amenable to normal significance testing, although 11 papers reported such tests. (See, for example, #s 13, 26, 39.)

[21] In several other papers, the authors do allude, directly or indirectly, to the economic significance or importance of their results. The justification, however, is attributed entirely to the result of a particular variable having tested statistically significant.

words, decisions made on the basis of the research would need to be addressed in the context of the relevant loss function.

**Randomness and Tests of Significance in Behavioral Studies**

An additional 11 articles in the eight journal issues we reviewed can be identified as behavioral accounting papers, and they are listed in Appendix B by letter. One of these papers also includes a regression analysis and, to this extent, is subject to the discussion above. Eight of these studies are based on samples of students. Of these eight, one is based on auditors working for the four largest auditing firms, a second examines experienced accountants, and the third is based on a "more representative" sample of participants taken from the American Mechanical Turk platform subject pool (see Appendix B, items H, J, and I, respectively). None of these studies is based on a random sample; hence, application of the probability calculus to an analysis of the sample data and statements of statistical significance is again only justified in the same sense as was discussed for the sample data used in the regression studies. In limited cases, reasonably complete listings of the possible sample data points may exist, which could closely approximate a population of interest and thus serve as a data source. For example, consider a list of actuaries or CPAs for a specific date and in a specified location. However, this approach is not used in our sample of papers and would be considered impracticable if not impossible in many cases of interest.

Behavioral laboratory studies have advantages over empirical studies when the issue at interest lends itself to a behavioral "laboratory" setting. In particular, a level of control can be achieved that is simply impossible in empirical research, relying on control variables. Falk and Heckman (2009) discuss these advantages in detail. For our purposes, however, these authors discuss three such factors that can create problems. They are the stakes or payments, participant experience, and the use of student subjects. It behooves the researcher to argue persuasively that the analysis is not limited by any of the three. Falk and Heckman (2009) document studies that led them to assert that the results are mixed to date on the question of whether the stakes or payments involved in the experiment affect the outcome. They further observe that participant experience can also be an issue muddying the interpretation of any results. But of the three, the use of student subjects is the most troublesome in attempting to generalize results. We first note that using student subjects is not a problem if the resulting predictions of the study can be shown to be independent of the subject pool. We conclude that researchers in behavioral accounting have a responsibility to address the extent to which each of the three concerns—and particularly the use of student subjects—has had an impact on the study's conclusions and whether they can be generalized. None of the behavioral studies in our sample meet this responsibility.

One of the more important advantages of behavioral studies, as we view them, is the opportunity, often present, to extend, repeat, or replicate a given study at a relatively low cost. In this way, any concern over stakes and experience might be explored. The participant issue is more difficult to resolve and is more costly. The low cost of replication, particularly in time, suggests research opportunities.

Of the 11 papers in our sample, six provide a discussion of the limitations and mention future research possibilities (see items A, B, F, I, J, K), while four others (see items C, D, E, H) do not address their study's limitations. And, with two notable exceptions (see items G and J), none of the other studies make any attempt to warn their readers of the inherent limitations implicit in the reported significance tests, nor otherwise justify why their results can be generalized to a population of interest.[22]

[22] One behavioral study (see item G) appropriately comments on an advantage of behavioral studies through employing a more controlled environment. The authors also point out the difficulty in generalizing the results to the larger population, however, leaving the matter for future research.

## SUGGESTIONS FOR FUTURE RESEARCH

Sampling of all kinds remains extensively used in the social sciences, including economics and accounting. But achieving the degree of sample randomness sufficient to allow the precise application of the probability calculus remains elusive if not impossible. We believe that the most important action that researchers can employ is to assure that important results in terms of economic impact are supported by replicating the research. Whether such replications will find acceptance in the various forms of publication remains to be seen. In the interim, we encourage the use of the SSRN and other Internet options.

In regard to first-time projects, we have explored how insights can be obtained from samples that do not fully qualify as random samples. When regression is the selected research approach, the researcher might consider combining it with analytics or with case analysis to lend support to the findings. We note that, in our sample, there are 11 regression studies listed in the SDF that relied in part on an analytical approach as well (see, for example, #s 22, 44, 49) and one that combined regression with a case approach (see # 40).

We further suggest increased reliance on the tools of descriptive statistics. The Ziliak and McCloskey (2004) paper is a good example. These approaches include graphs, tables, and especially ratios of the actual observations to provide insights into the questions of interest. In doing so, it is imperative to keep in mind that, while ratios of actual economic phenomena are not probabilities, they can be extremely useful in comparative analysis.

An example is the use of fractiles, which avoid the assumption of a middle point, other than a median, from which data are treated as deviating randomly. Fractiles have the additional advantage that inter-fractile distances can be used to measure variability. We observe that about slightly over 50 percent of the studies report fractile data along with their sample's descriptive statistics (see #s 4, 29, 39), and 55 percent provided graphs, charts, or tables to support their work and help readers visualize the findings (see #s 14, 40, 53, and the SDF). We encourage more of this type of presentation and analysis when using fractiles. However, we also urge careful consideration of when important patterns are better perceived through graphs or charts rather than in tables.

Using graphical techniques is another way to illustrate and examine relationships among important subsets of variables when a large number of independent variables are used. An interesting exercise would be to select several pairs of independent variables, randomly or purposely, from the data set and plot the three-dimensional relationships with the dependent variable. One result would be a visual of the limited relationships, which the analysis assumes, perhaps incorrectly, are essentially linear after the applied variable transformations.

We have, however, not been able to advance a parsimonious means of addressing accounting processes involving many explanatory variables that can duplicate insights easily obtainable using computer programs. Therefore, until improved techniques emerge, we expect to see continued use of regression analysis but accompanied by explicit recognition of what it can and cannot do as well as what it did and did not do.[23]

Perhaps a more constructive approach would be to develop alternative approaches to regression analysis that would help explain interesting issues, while simultaneously requiring less in the realm of limiting assumptions.[24]

---

[23] One technique used in management accounting, particularly budgeting, is moderated regression (see Hartmann and Moers 1999, 2003). This technique can be useful when theory posits interaction effects. Johnstone (1997) and Ioannidis (2005) address the approach to regression studies by using Bayesian statistics.

[24] See Winkler (2009), particularly Chapters 4, 8, and 9. See also section 9.5 and Appendix D for an interesting (but limited) regression alternative that avoids the limitations of traditional NHSTs.

## "CONCLUSIONS" SECTIONS: AUTHORS' DISCUSSIONS OF LIMITATIONS AND DIRECTIONS FOR FURTHER RESEARCH

The large majority of the discussion in the "Conclusions" section for most of the papers we reviewed is generally confined to a discussion of the research's incremental contribution to the literature and why it should be of interest to the reader. The Conclusions section should also provide a discussion of the study's limitations, including what might, if anything, be done about them. The Conclusions section is also an ideal place to suggest the possibilities for future research and even suggestions for how to proceed. The authors are in an excellent position to address these issues, and some do comment on them as they arise in their studies. The majority of the papers address at least one of these issues.[25] A few papers do a creditable job (see, for example, #s 16, 24, 34, and the SDF). However, 42 percent of the articles provide no final discussion of their study's limitations or future research opportunities in either a separate section or as part of their conclusions (see, for example, #s 19, 43, 46, and the SDF).

We are impressed by the thoroughness of the attention given by the authors to those who have gone before them. However, as indicated above, we are troubled that the studies relied upon have seldom, if ever, been replicated. Replication helps establish what is correct and what is fraudulent. We have elected not to frame our paper as an effort to detect fraud. While fraud is not unknown in either the social sciences or physical scientific research, it is seldom acknowledged and extraordinarily difficult to detect.[26]

The research methods used today have been developed by those who have gone before, but justifying these methods today appears to depend only on whether the approach appears applicable to the data. This situation only emphasizes the need for replication, even if the likelihood of publication of replication research remains slim, at best, for now. Packaging replications with new or extended analysis may provide an avenue for publication. Modifying the academic award structure, as difficult as that may be, would also be helpful.

We invite, indeed encourage, replications of our study for this or other sets of journal articles to determine whether the issues raised here are found elsewhere in the accounting literature or, more importantly, will be present in the future. A replication using articles in other accounting journals over the same period would be an example. The working paper by Leone et al. (2014), referenced earlier, provides a set of 857 accounting papers for the period 2006–2010. We expect the issues examined in our paper to be apparent in other current accounting journals as well as in prior research even though analysis processes have improved over time. Perhaps, for example, had there been more replications examining the implicit assumptions underlying prior work, much referenced in all of the papers we have examined, progress would have been even more pronounced, resulting in the recognition and avoidance of the issues raised here.

## IMPLICATIONS FOR AUTHORS, THEIR READERS, EDITORS, REVIEWERS, AND EDUCATORS

In summary we have endeavored to make the following points:

---

[25] We admit that the decision not to address future research opportunities may simply reflect the authors' view that such opportunities are already suggested by limitations noted earlier in their article.

[26] The existence of fraud in the sciences, but not in accounting, is identified by Broad and Wade (1982). See, in particular, Chapter 2 in which frauds committed by some of the most illustrious figures in history are described, Chapter 8 for a recent famous case which unfolded at Cornell University, and, finally, the Appendix, where 168 known or suspected cases of fraud are recorded. See also Bailey et al. (2001) for an interesting and revealing documentation of the frequency of fraud in accounting studies.

First, authors must adequately defend their selection of the sample period by convincing the reader that the period is stable itself and in relation to periods in close proximity.

Second, the accounting academy should actively seek and reward replications as an essential element in its aspirations to be a scientific community.

Third, authors should attend to the economic significance as well as the statistical significance of their investigations.

Fourth, authors should respect the limitation of conventional hypothesis tests applied to their data, which implies enhanced caution when declaring results to be statistically significant.

Fifth, authors could consider reporting the use of statistical intervals as a way to mitigate the problems of determining the most likely alternative hypothesis and thereby the appropriate Type II error.

Sixth, authors need to be sure that, in their "Conclusions" section, they discuss the limitations of their research and how these limitations might be overcome, as well as suggest extensions for future research.

Seventh, authors should consider the use of descriptive statistics and other approaches as a means of, or support for, establishing the validity of their research objective.

Eighth, editors should consider requiring authors of accepted papers to provide a complete description of their methodology, including data collection, accuracy, and verification procedures, as well as to have them sign a formal agreement to retain a complete file for a specified period.

Ninth, authors should be required to submit the results of all tests performed and the data used and rejected, the latter with justifications, for each test. The data should be supplied in a form suitable for uploading into a standard or supplied statistical package and include the appropriate software commands.

Tenth, editors should assure that any empirical paper is reviewed by a competent statistician, and educators in doctoral programs should ensure that their graduates are properly prepared in research methodology and statistics.

Our fervent hope is that this paper will engender a discussion of how to improve research in all areas of accounting. It appears to be needed. We welcome those who would challenge our observations, analyses, and recommendations. Let the discussion continue.

## REFERENCES

Atiase, R. K. 1985. Predisclosure information, firm capitalization, and security price behavior around earnings announcements. *Journal of Accounting Research* 23 (1): 21–36.

Bailey, C. D., J. R. Hasselback, and J. N. Karcher. 2001. Research misconduct in accounting literature: A survey of the most prolific researchers' actions and beliefs. *Abacus* 37 (1): 6–54.

Ball, R., and P. Brown. 1968. An empirical evaluation of accounting income numbers. *Journal of Accounting Research* 6 (2): 159–178.

Balsam, S., A. J. Irani, and Q. J. Yin. 2012. Impact of job complexity and performance on CFO compensation. *Accounting Horizons* 26 (3): 395–416.

Bamber, L. S., T. E. Christensen, and K. M. Gaver. 2000. Do we really "know" what we think we know? A case study of seminal research and its subsequent overgeneralization. *Accounting, Organizations and Society* 25 (2): 103–129.

Beaver, W. H. 1968. The information content of annual earnings announcements. *Empirical Research in Accounting, Selected Studies 1968*, Supplement to *Journal of Accounting Research* 6: 67–92.

Begley, C. G., and L. M. Ellis. 2012. Drug development: Raise standards for preclinical cancer research. *Nature* 483 (March 29): 531–533.

Broad, W., and N. Wade. 1982. *Betrayers of the Truth*. New York, NY: Simon and Schuster.

Brown, P. R. 2013. How can we do better? *Accounting Horizons* 27 (4): 855–859.
Chambers, R. J. 1973. Observation as a method of inquiry—The background of *Securities and Obscurities*. *Abacus* 9 (2): 156–175.
Chambers, R. J. 1974. Stock market prices and accounting research. *Abacus* 10 (1): 39–54.
Chow, C. W. 1983. The impacts of regulation on bondholder and shareholder wealth: The case of the Securities Acts. *The Accounting Review* 58 (3): 485–520.
Cumming, G. 2013. The new statistics: How and why. *Psychological Science* OnlineFirst (November 12): 1–24.
Dyckman, T. R., D. H. Downes, and R. P. Magee. 1975. *Efficient Capital Markets and Accounting: A Critical Analysis*. Englewood Cliffs, NJ: Prentice Hall.
*Economist, The*. 2013. Unreliable research: Trouble at the lab. U.S. Edition. (October 19): 26–30.
*Economist, The*. 2014. Combating bad science: Metaphysicians. U.S. Edition. (March 15): 74.
Eich, E. 2013. Business not as usual. Editorial. *Psychological Science* OnlineFirst (November 27): 1–4.
Falk, A., and J. J. Heckman. 2009. Lab experiments are a major source of knowledge in the social sciences. *Science* October 23: 535–538.
Hamermesh, D. S. 2000. The craft of labormetrics. *Industrial and Labor Relations Review* 53 (3): 363–380.
Hartmann, F. G. H., and F. Moers. 1999. Testing contingency hypotheses in budgetary research: An evaluation of the use of moderated regression analysis. *Accounting, Organizations and Society* 24 (4): 291–315.
Hartmann, F. G. H., and F. Moers. 2003. Testing contingency hypotheses in budgetary research using moderated regression analysis: A second look. *Accounting, Organizations and Society* 29 (7-8): 803–809.
Ioannidis, J. P. A. 2005. Why most published research findings are false. *PLoS Medicine* 2 (8): 96–701.
Jasny, B. R., G. Chin, L. Chong, and S. Vignieri. 2011. Again, again, and again. . . *Science* (December 2) 334: 1225.
Johnson, V. 2013. Uniformly most positive Bayesian tests. *Annals of Statistics* 41 (4): 1716–1741.
Johnstone, D. J. 1988. Comments on Oakes on the foundations of statistical inference in the social and behavioral sciences: The market for statistical significance. *Psychological Reports* 63 (1): 319–331.
Johnstone, D. J. 1997. Comparative classical and Bayesian interpretations of statistical compliance tests in auditing. *Accounting and Business Research* 28 (1): 53–82.
Kuhn, T. S. 1962. *The Structure of Scientific Revolutions*. Chicago IL: The University of Chicago Press.
Kyburg, H. E. 1976. Chance. *Journal of Philosophical Logic* 5: 355–393.
Leone, A., M. Minutti-Meza, and C. Wasley. 2014. *Outliers and Inference in Accounting Research*. Working paper, University of Miami, School of Business.
Martin, A., and G. Naik. 2014. Stem-cell finding criticized. *The Wall Street Journal* (March 11): 3.
Merino, B. D., B. S. Koch, and K. L. MacRitchie. 1987. Historical analysis—A diagnostic tool for "event" studies: The impact of the Securities Act of 1933. *The Accounting Review* 62 (4): 748–762.
Naik, G., and A. Martin. 2014. Scientists struggle to reproduce new stem-cell finding. *The Wall Street Journal* (February 20). Available at: http://online.wsj.com/news/articles/SB10001424052702303636404579392551374908642
Obokata, H., T. Wakayama, Y. Sasai, K. Kojima, M. P. Vacanti, H. Niwa, M. Yamato, and C. A. Vacanti. 2014. Stimulus-triggered fate conversion of somatic cells into pluripotency. *Nature* 505 (January 30): 641–647.
Obokata, H., Y. Sasai, H. Niwa, M. Kadota, M. Andrabi, N. Takata, M. Tokoro, Y. Terashita, S. Yonamura, C. A. Vacanti, and T. Wakayama. 2014. Bidirectional developmental potential in reprogrammed cells with acquired pluripotency. *Nature* 505 (January 30): 676–680.
Oppong, A. 1980. Information content of annual earnings announcements revisited. *Journal of Accounting Research* 18 (2): 574–584.
Taleb, N. N. 2007. *The Black Swan: The Impact of the Highly Improbable*. New York, NY: Random House.
Taleb, N. N. 2012. *Antifragile: Things That Gain from Disorder*. New York, NY: Random House.
Watts, R. L., and J. L. Zimmerman. 1986. *Positive Accounting Theory*. New York, NY: Prentice Hall.
Winkler, O. 2009. *Interpreting Economic and Social Data*. Dordrecht, The Netherlands: Springer.

Ziliak, S. T., and D. N. McCloskey. 2004. Size matters: The standard error of regressions in the American Economic Review. *The Journal of Socio-Economics* 33 (5): 527–546.

## APPENDIX A

## List of the Fifty-Five Articles in the Sample

***The Accounting Review*—September 2012**

1. Blouin, Krull, and Robinson—Is U.S. Multinational Dividend Repatriation Policy Influenced by Reporting Incentives?
2. Cheng, Huang, Li, and Stanfield—The Effect of Hedge Fund Activism on Corporate Tax Avoidance
3. Crawford, Roulstone, and So—Analyst Initiations of Coverage and Stock Return Synchronicity
4. Du, Tang, and Young—Influence Activities and Favoritism in Subjective Performance Evaluation: Evidence from Chinese State-Owned Enterprises
5. Hoopes, Mescall, and Pittman—Do IRS Audits Deter Corporate Tax Avoidance?
6. Keune and Johnstone—Materiality Judgments and the Resolution of Detected Misstatements: The Role of Managers, Auditors, and Audit Committees
7. Roychowdhury and Sletten—Voluntary Disclosure Incentives and Earnings Informativeness
8. Shanthikumar—Consecutive Earnings Surprises: Small and Large Trader Reactions
9. Skinner and Srinivasan—Audit Quality and Auditor Reputation: Evidence from Japan
10. Yip and Young—Does Mandatory IFRS Adoption Improve Information Comparability?
11. Zhang—The Effect of *Ex Ante* Management Forecast Accuracy on the Post-Earnings-Announcement Drift

***The Accounting Review*—November 2012**

12. Atwood, Drake, Myers, and Myers—Home Country Tax System Characteristics and Corporate Tax Avoidance: International Evidence
13. Burnett, Cripe, Martin, and McAllister—Audit Quality and the Trade-Off between Accretive Stock Repurchases and Accrual-Based Earnings Management
14. Dechow and You—Analysts' Motives for Rounding EPS Forecasts
15. Donelson, McInnis, Mergenthaler, and Yu—The Timeliness of Bad Earnings News and Litigation Risk
16. Florou and Pope—Mandatory IFRS Adoption and Institutional Investment Decisions
17. Gordon and Wilford—An Analysis of Multiple Consecutive Years of Material Weaknesses in Internal Control
18. Kim, Liu, and Zheng—The Impact of Mandatory IFRS Adoption on Audit Fees: Theory and Evidence
19. Lee, Matsunaga, and Park—Management Forecast Accuracy and CEO Turnover
20. Liu and Natarajan—The Effect of Financial Analysts' Strategic Behavior on Analysts' Forecast Dispersion
21. Skantz—CEO Pay, Managerial Power, and SFAS 123(R)

***The Accounting Review*—January 2013**

22. Banker, Darrough, Huang, and Plehn-Dujowich—The Relation Between CEO Compensation and Past Performance

23. Efendi, Ouyang, and Swanson—Executive Turnover Following Option Backdating Allegations
24. Erkens and Bonner—The Role of Firm Status in Appointments of Accounting Financial Experts to Audit Committees
25. Gu, Li, and Yang—Monitors or Predators: The Influence of Institutional Investors on Sell-Side Analysts
26. Iskandar-Datta and Jia—Valuation Consequences of Clawback Provisions
27. Kaplan and Williams—Do Going Concern Audit Reports Protect Auditors from Litigation? A Simultaneous Equations Approach
28. Kilic, Lobo, Ranasinghe, and Sivaramakrishnan—The Impact of SFAS 133 on Income Smoothing by Banks through Loan Loss Provisions
29. Naiker, Navissi, and Truong—Options Trading and the Cost of Equity Capital
30. Naiker, Sharma, and Sharma—Do Former Audit Firm Partners on Audit Committees Procure Greater Nonaudit Services from the Auditor?

***The Accounting Review*—March 2013**

31. Chen, Martin, and Wang—Insider Trading, Litigation Concerns, and Auditor Going-Concern Opinions
32. Chuk—Economic Consequences of Mandated Accounting Disclosures: Evidence from Pension Accounting Standards
33. De George, Ferguson, and Spear—How Much Does IFRS Cost? IFRS Adoption and Audit Fees
34. Demerjian, Lev, Lewis, and McVay—Managerial Ability and Earnings Quality
35. Francis and Michas—The Contagion Effect of Low-Quality Audits
36. Joos and Leung—Investor Perceptions of Potential IFRS Adoption in the United States
37. Kecskés, Mansi, and Zhang—Are Short Sellers Informed? Evidence from the Bond Market
38. Mola, Rau, and Khorana—Is There Life After the Complete Loss of Analyst Coverage?

***Journal of Accounting Research*—September 2012**

39. Bushman and Wittenberg-Moerman—The Role of Bank Reputation in "Certifying" Future Performance Implications of Borrowers' Accounting Numbers
40. Campbell—Employee Selection as a Control System
41. Chen and Sandino—Can Wages Buy Honesty? The Relationship Between Relative Wages and Employee Theft
42. Drake, Roulstone, and Thornock—Investor Information Demand: Evidence from Google Searches Around Earnings Announcement
43. Ozkan, Singer, and You—Mandatory IFRS Adoption and the Contractual Usefulness of Accounting Information in Executive Compensation

***Journal of Accounting Research*—December 2012**

44. Ammer, Holland, Smith, and Warnock—U.S. International Equity Investment
45. Hutton, Lee, and Shu—Do Managers Always Know Better? The Relative Accuracy of Management and Analyst Forecasts
46. Klassen and Laplante—Are U.S. Multinational Corporations Becoming More Aggressive Income Shifters?
47. Lui, Markov, and Tamayo—Equity Analysts and the Market's Assessment of Risk

***Journal of Accounting Research*—March 2013**

48. Ahmed and Duellman—Managerial Overconfidence and Accounting Conservatism
49. Ball—Does Anticipated Information Impose a Cost on Risk-Averse Investors? A Test of the Hirshleifer Effect
50. Dhaliwal, Kaplan, Laux, and Weisbrod—The Information Content of Tax Expense for Firms Reporting Losses
51. Firth, Lin, Liu, and Xuan—The Client is King: Do Mutual Fund Relationships Bias Analyst Recommendations?
52. Kama and Weis—Do Earnings Targets and Managerial Incentives Affect Sticky Costs?

***Journal of Accounting Research*—May 2013**

53. Bonsall, Bozanic, and Fischer—What Do Management Earnings Forecasts Convey About the Macroeconomy?
54. Knechel, Niemi, and Zerni—Empirical Evidence on the Implicit Determinants of Compensation in Big 4 Audit Partnerships
55. Li, Lundholm, and Minnis—A Measure of Competition Based on 10-K Filings

## APPENDIX B

## Sample of Eleven Behavioral Articles

***The Accounting Review*—November 2012**

A. Chen, Williamson, and Zhou—Reward System Design and Group Creativity: An Experimental Investigation
B. Christ, Sedatole, and Towry—Sticks and Carrots: The Effect of Contract Frame on Effort in Incomplete Contracts

***The Accounting Review*—January 2013**

C. Bennett and Hatfield—The Effect of Social Mismatch between Staff Auditors and Client Management on the Collection of Audit Evidence
D. Davidson and Stevens—Can a Code of Ethics Improve Manager Behavior and Investor Confidence? An Experimental Study
E. Tafkov—Private and Public Relative Performance Information under Different Compensation Contracts

***The Accounting Review*—March 2013**

F. Falsetta, Rupert, and Wright—The Effect of the Timing and Direction of Capital Gain Tax Changes on Investment in Risky Assets
G. Hannan, McPhee, Newman, and Tafkov—The Effect of Relative Performance Information on Performance and Effort Allocation in a Multi-Task Environment
H. Libby and Brown—Financial Statement Disaggregation Decisions and Auditors' Tolerance for Misstatement

***Journal of Accounting Research*—December 2012**

I. Rennekamp—Processing Fluency and Investors' Reactions to Disclosure Readability

***Journal of Accounting Research*—March 2013**

J. Chen, Tan, and Wang—Fair Value Accounting and Managers' Hedging Decisions
K. Choi, Hecht, and Taylor—Strategy Selection, Surrogation, and Strategic Performance Measurement Systems

## APPENDIX C

Appendix_A_B: http://dx.doi.org/10.2308/acch-50818.s1