



PERGAMON

Accounting, Organizations and Society 25 (2000) 103–129

Accounting,
Organizations
and Society

www.elsevier.com/locate/aos

Do we really ‘know’ what we think we know? A case study of seminal research and its subsequent overgeneralization

Linda Smith Bamber^{a,*}, Theodore E. Christensen^b, Kenneth M. Gaver^a

^a*J.M. Tull School of Accounting, 255 Brooks Hall, University of Georgia, Athens, GA 30602-6252, USA*

^b*Department of Accountancy, Weatherhead School of Management, Case Western Reserve University, Cleveland, OH 44106-7235, USA*

Abstract

We show that the community of accounting researchers has not fully appreciated the sensitivity of research conclusions to (necessarily) subjective research design choices, and that this failure has led to the subsequent overgeneralization of early evidence. Our analysis is based on a case study that examines the effects of two basic research design choices, and subsequent researchers’ appreciation of the impact of those choices. To enhance our study’s accessibility to a broad cross-section of the accounting research community, our case study focuses on an article that is covered in most accounting doctoral programs — the recipient of the second American Accounting Association “Seminal Contribution to Accounting Literature Award” — Beaver’s 1968 article on the information content of annual earnings announcements [Beaver, W. (1968). The information content of annual earnings announcements. *Empirical research in accounting; Selected studies, 1968*. Supplement to *Journal of Accounting Research*, 6, 67–92.]

Our analysis reveals that the research design choices in Beaver’s study significantly affected the results. Beaver clearly explained these choices, and was careful in drawing inferences. Nevertheless, *subsequent researchers* interpreted Beaver’s evidence too broadly. Our new empirical evidence suggests that the information content inference is much more fragile than is generally suspected. We also present citation analyses showing that, consistent with the Kuhnian view that adherents of a paradigm tend to ignore (later) anomalous evidence, subsequent researchers largely overlooked post-Beaver evidence that was inconsistent with the paradigm’s interpretation of Beaver’s results. The paper concludes with a brief consideration of how cognitive biases of individuals, combined with biases inherent in the review process and the academy in general, foster an environment where the placement of the first research bricks affect the whole wall. While the case study focuses on an archival empirical financial accounting study, consideration of the effects of subjective research design choices and the documented deficiencies in the interpretation of research are relevant to empirical accounting researchers in general and to critical theorists, as well.

While accounting scholars ‘think we understand’ the impact of research design choices, the evidence presented here suggests that we have collectively failed to apply this understanding in practice. Such failure delays the acquisition of knowledge. This study provides a salient demonstration indicating that we must constantly remain on guard to recognize the subjective nature of research design choices, to consider the likely effects of those choices, and to be cautious in drawing generalizations. © 2000 Elsevier Science Ltd. All rights reserved.

* Corresponding author. Tel.: +1-706-542-3501; fax: +1-706-542-3630.

E-mail address: lbamber@terry.uga.edu (L. Smith Bamber).

1. Introduction

What is a fact? A fact is merely an opinion held by all. . . Leamer (1983).

This paper examines the development of a major accounting research paradigm in the context of Kuhn's (1970) scientific revolution framework. Consistent with Kuhn's arguments that a paradigm's disciples take for granted their paradigm's foundations, we show that accounting researchers have failed to appreciate fully the likely effects of (necessarily) subjective research design choices made in seminal research. Our study shows that research design choices significantly affected evidence reported in early research. We also show that subsequent research generalized initial seminal studies' results too broadly, and largely overlooked subsequent evidence based on complementary research design choices that reported results contrary to what had become the paradigm's core beliefs. Thus, while accounting scholars 'think we understand' the impact of subjective research design choices, the evidence presented here suggests that we have collectively failed to apply this understanding in practice. Such failures are unfortunate because they delay the acquisition of knowledge.

This issue is important in all areas of accounting research, and we address this void in the literature by providing new empirical evidence on the impact of research design choices in the context of an archival empirical study that should be familiar to most accounting scholars: Beaver's (1968) seminal paper, *The Information Content of Annual Earnings Announcements*. This is the second of only two papers to receive the American Accounting Association's prestigious "Seminal Contribution to Accounting Literature Award,"¹ and Brown (1996) identifies Beaver (1968) as one of the ten most influential articles in accounting research. At the outset, we emphasize that our purpose is *not* to criticize Beaver's seminal study or diminish its importance as a significant contribution to the literature. Rather, *our criticism focusses on the subsequent*

behavior of the academy in interpreting and propagating the implications of Beaver's study.

While the case study focusses on Beaver (1968), our study contributes more broadly to an understanding of the conduct and interpretation of research. We focus on often overlooked effects of subjective research design choices, and we document deficiencies in *subsequent researchers'* interpretation of seminal research. These issues are as relevant to experimental accounting researchers and critical theorists as they are to financial accounting researchers.

Our analysis begins with a review of the historical context in which Beaver (1968) was conducted and published. This historical context is key to understanding why Beaver's study became so influential, and in particular, why it has become the basis upon which subsequent research inferred the 'stylized fact' that earnings announcements convey new information to the market.² First, consistent with Kuhn's (1970) view that adherents of a paradigm tend to ignore (later) anomalous evidence, we show that subsequent researchers largely overlooked post-Beaver evidence that was inconsistent with the paradigm's interpretation of Beaver's results. Second, while Beaver was careful not to overstate his study's results, our citation analysis shows that subsequent research did interpret Beaver's evidence too broadly, consistent with Kuhn's (1970) claim that later disciples of a paradigm take for granted the paradigm's foundations. Specifically, subsequent researchers largely failed to recognize the impact of two aspects of Beaver's research design choices on the study's empirical results. Subsequent research has implicitly inferred that the mean result emphasized in the study is 'typical' of market reactions to individual earnings announcements. And subsequent research has also failed to appreciate the implications of Beaver's sample selection criteria for the empirical

² Like Beaver (1968), we focus on the incremental information content of accounting earnings announcements — whether the announcements convey "new news" to the market, as indicated by unusual price movements. Incremental information content requires not only that accounting earnings announcements convey value-relevant information, but also that these accounting disclosures are timely enough to convey new value-relevant information to the market.

¹ Ball and Brown (1968) was the first recipient of the award.

results. After considering the historical evolution of the information content research paradigm, we provide new empirical evidence on the impact of each of these research design choices. Our evidence suggests the inferences subsequent research drew from Beaver's (1968) results are more fragile than is generally suspected.

With respect to the first research design choice — the emphasis on the cross-sectional mean market reaction — we find that *a focus on mean effects obscures the fact that most individual earnings announcements are not associated with unusual price reactions*. Thus, the first misinterpretation of Beaver's evidence is the tendency to view mean market reactions (such as those saliently presented in Beaver's famous Figs. 1, 3 and 6) as representative of the *typical* stock price reactions to individual earnings announcements. This misinterpretation can be viewed as a manifestation of Kahneman and Tversky's (1972) representativeness heuristic. As Hogarth explains, "We tend to imagine that what we see... is *typical* of what can occur... Hence we tend to assess distributions which are too 'tight'." (Hogarth, 1987, p. 268, emphasis in original).

With respect to the second research design choice — sample selection criteria — subsequent research citing Beaver's information content result has largely overlooked the fact that Beaver's sample was based on a set of relatively small firms.³ Beaver's sample included only non-12/31 NYSE firms with less than 20 news items in the *Wall Street Journal* (*WSJ*, hereafter) during that calendar year. These sample selection criteria led to a sample of 143 relatively small firms that comprised only about 5% of the period's total market value of NYSE firms. To assess the effects of these sample selection criteria, we repeat our analyses, choosing an alternative set of firms that would have been at least equally interesting at the time when there was no empirical evidence on the information content of any firms' earnings announcements — the era's Fortune 200 firms. As

these firms comprised 67% of the market value of the NYSE during that time period, the information content of these earnings announcements would have been of interest in their own right. With 30 years of research hindsight, it is not surprising to find that the Fortune 200 firms' earnings announcements generate smaller market reactions than do earnings announcements of 'Beaver-like' firms. What is more surprising is that for Fortune 200 firms, the announcement week mean price reaction is not statistically different from the mean price movements in 11 of the 16 surrounding weeks. Further, *most* of the individual Fortune 200 earnings announcements do not generate unusual market reactions.

The final section of this paper argues that cognitive biases of individuals combined with biases in the review process and the academy in general foster an environment where the "placement of the first [research] bricks affect the whole wall." Had the initial information content studies made different research design choices — either by focusing on the extent to which *individual earnings announcements* were informative, or by not using samples of relatively small firms, the apparent information content would have been much less dramatic, and perhaps even non-existent. However, the research design choices made in Beaver (1968) and other early information content studies (e.g. May, 1971) yielded results suggesting that earnings announcements had extensive information content. Although Beaver was careful not to over-generalize his results, the conclusion that earnings announcements convey new information to the market soon became the received wisdom.

In the Kuhnian sense, subsequent researchers took this notion for granted as one of the pillars of the price-earnings research paradigm: "Any person familiar with the 'real world' knows that the market reacts strongly (and perhaps overreacts) to earnings-per-share data" (Kaplan, 1978, p. 138). This view was not seriously challenged for many years, thereby delaying the acquisition of knowledge. It was not until 17 years after Beaver (1968) that Atiase's (1985) seminal investigation of cross-sectional differences in the information content of earnings announcements revealed that small firms' announcements are more informative

³ It has also been largely forgotten that the next major information content study, May (1971), examined the market's reaction to interim earnings announcements of AMEX firms, which are also smaller on average than NYSE firms.

than large firms' announcements, on average. Now our study presents new evidence that most earnings announcements — even those of the small 'Beaver-like' firms — are not associated with unusual price reactions. This evidence raises questions about the information content of 'typical' earnings announcements, which in turn supports the Jenkins Committee's call for expansion of the role of accounting and auditing beyond the conventional financial statements to encompass a broader notion of business reporting (American Institute of Certified Public Accountants, AICPA, 1994).

In sum, we show that research design choices significantly affected the results of seminal information content studies, but subsequent research largely failed to recognize the impact of these choices on the generalizability of the early evidence. While accounting scholars "think we understand" that research design choices affect research results and their generalizability, the evidence presented here suggests that we have collectively failed to apply this understanding in practice. This delays the acquisition of knowledge and results in misallocation of research resources (for example, decades passed before serious questions were raised about the extent to which earnings announcements convey 'new news' to the market).⁴ The intent of this study is to reduce our collective susceptibility to such biases in the future, by providing a salient demonstration why we must constantly remain on guard to recognize the subjective nature of research design choices, to consider the likely effects of those choices, and to be cautious in drawing generalizations.

⁴ Mouck (1998) also raises questions about the validity of inferences from capital markets research in accounting, but his questions are based on the assumption that security prices are not reasonable first-order approximations of firm value. Our study demonstrates that even without Mouck's assumption — that is, even within the existing capital markets research paradigm — one can still raise serious questions about certain inferences that have become embedded in the core foundation of the price-earnings research paradigm.

⁵ Wells (1976) and Mouck (1993) also discuss accounting research in the context of Kuhn's (1970) theory of scientific revolutions. In contrast to those studies, we focus on the role Beaver (1968) played in the development of modern price-earnings research, viewed in the context of Kuhn's framework.

2. Evolution of thought regarding the information content of accounting earnings announcements

2.1. *Development of the paradigm*

Appreciation of the importance of Beaver's (1968) seminal evidence and the effect of misinterpretation of his evidence by subsequent research requires an understanding of the historical context in which Beaver's study was conducted and published. We frame the development of accounting research in terms of Kuhn's (1970) description of the evolution from unorganized, unstructured, independent "pre-paradigm" activity toward the development of a scientific paradigm and the practice of what Kuhn terms "normal science."^{5,6} We describe the pivotal role Beaver played in stimulating the development of the price-earnings research paradigm. We also show that subsequent adherents of the paradigm largely overlooked anomalous evidence counter to the paradigm's core beliefs, as Kuhn suggests is typical in normal science.

2.1.1. *Pre-paradigm activity in accounting*

Influential writers like Paton and Littleton (1940, pp. 98–99) advanced the so-called "entity theory" view that corporate earnings is the focal point of accounting.⁷

⁶ In relying on Kuhn's framework, we also use his labels for the stages of scientific evolution: (1) 'pre-paradigm' activity, (2) a paradigm-shifting 'notable scientific achievement,' and (3) 'normal science.' Kuhn concludes that normal science is fraught with shortcomings, including a tendency to dismiss anomalous evidence and to focus on increasingly trivial research issues (which he terms 'puzzle-solving'). We provide evidence supporting these allegations in the context of accounting research. Thus, the label 'normal science' should not be construed as a synonym for 'truth-discovery' and more generally, the labels 'pre-paradigm' and 'normal science' do not imply the superiority of one over the other.

⁷ In contrast, recent FASB pronouncements tend to emphasize the primacy of the balance sheet and the appropriate valuation of assets and liabilities, rather than the income statement (e.g., SFAS 109 changed accounting for deferred income taxes from the deferred method to the liability method). We thank former FASB chairman Dennis Beresford for pointing this out to us.

[B]usiness men and public accountants have been placing an increasing emphasis on the income statement in recent years—[such] that the income report has supplanted the balance sheet as a focus of attention.

This perspective suggests that the definition and computation of corporate earnings, and the resulting informativeness of those earnings, are among the most critical issues facing accountants.

Consequently, it is not surprising that the accounting literature prior to 1968 was dominated by normative position papers centering on the definition and computation of corporate earnings — the nature of ‘true economic income’, how this construct should be measured, and how close the day’s accounting rules were to reporting ‘true economic income’. Since no common set of beliefs or methods existed, each writer laid arguments, starting anew from independent and study-specific foundations (American Accounting Association, 1977). This led to an unorganized and unstructured body of literature with divergent assumptions and conclusions, and no apparent means of achieving consensus. Mouck (1993) points out that this is characteristic of what Kuhn terms “pre-paradigm” activity.

Some authors of the period suggested that investors find accounting earnings useful in making investment decisions.

As long ago as September 1932, a committee of the [American] Institute [of Accountants] said, ‘... the income account is usually far more important than the balance sheet.’ This shift in the emphasis placed on financial statements is largely due to the significance that has been placed on earnings capacity in valuing corporate securities (Blough, 1937, p. 283).

Investors use accountants’ reported income as an index, for when a past income increases, investors interpret it to mean that the index of enterprise success has gone up. They therefore buy the stock at a higher price.... (Bedford, 1957, p. 60).

[T]he annual net earnings figure tends to have a magical significance — not only for the

ordinary investor but for security analysts and even for acquisition-minded managements (*Forbes*, May 1967, p. 28).

Other authors scoffed at the usefulness of earnings, arguing that the arbitrary nature of accounting procedures and the failure to require their uniform application led to ‘no very accurate measure of the profits of corporations’. (Greer, 1938, p. 341).

[N]et income can never be supposed to be a fact in any sense at all except that it is the figure that results when the accountant has finished applying the procedures which he adopts (Canning, 1929, p. 99).

[One author asserts that] the accepted method of making up profit and loss statements is remarkable for its obscuring of its essential facts....[another] author compares the traditional methods of valuation to ‘measuring a field with a rubber tape line’ (Gilman 1939, p. 5).

Entity income is not a meaningful concept.... [N]or can entity [income] calculations be readily translated into meaningful data for the investor, so far as the orthodox form of income report is concerned. (Vatter, 1947, pp. 33–34).

All invented values, such as L.I.F.O. valuations, calculated depreciation figures, amounts of ‘deferred’ charges and credits, were not measures of anything. (Chambers, 1974, p. xiii).

Some opined that even if earnings were meaningful, the potential information content of earnings *announcements* would be usurped by more timely sources of information, and accounting earnings announcements would be too late to be useful to investors:

I would argue that reported earnings are of key importance in investment decisions but that they become gradually more deducible and, in fact, are generally relatively accurately deduced before they are reported (Parker, 1967, p. 16).

Thus, pre-1968 discussion of the informativeness of earnings either supported the utility of earnings for investors, or else expressed skepticism concerning the informativeness of earnings. This dialogue aptly illustrates Kuhn's description of pre-paradigm activity:

in the early stages of the development of any science different men confronting the same range of phenomena...describe and interpret them in different ways (Kuhn, 1970, p. 17).

While the normative position papers provided important bases for development of the accounting discipline (including identification of corporate earnings as a legitimate issue for study), the inductive and deductive reasoning these papers employed was incapable of resolving their differences. In the absence of empirical evidence, the discipline was not converging on the important question whether accounting earnings announcements *do* (as opposed to *should*) convey 'new news' to investors.

2.1.2. "Notable scientific achievements" leading to the birth of the price-earnings research paradigm

During the 1960's, advances in computational power permitted collection and analysis of data on a much larger scale than was previously possible. The coincident development of the Efficient Markets Hypothesis, the Capital Asset Pricing Model,

and event study methodology,⁸ along with the availability of massive quantities of security markets data in machine-readable form (from CRSP), set the stage for large-scale empirical research that could investigate the price-earnings relation. The early capital market studies in accounting (Ball & Brown 1968; Beaver 1968; Benston 1967) provided the "notable scientific achievements" that Kuhn (1970, p. 178) suggests are necessary for the establishment of a scientific paradigm.

One smaller-scale study of the earnings-price relation had already appeared, serving as a precursor to the more sophisticated, larger-scale, late-1960's studies that would be perceived as notable scientific achievements. Ashley (1962) reported that extremely large earnings increases (between 100 and 200% or over 200%) were on average associated with price increases around the earnings announcement date. Conversely, steep earnings drops (between 25 and 50% or over 50%) were associated with declining prices. Although these results should have been of interest to the new breed of accounting researchers, the Social Sciences Citation Index reveals only four citations of Ashley's information content result in the accounting literature (and two of these citations are by the same set of authors). Thus, accounting researchers behaved as if they were unaware of empirical evidence on the relation between corporate earnings and prices.⁹

⁸ Ball (1994) discusses the critical importance of the development of event study methodology. He notes that even as late as the mid-1960's the "stock market was generally viewed as exhibiting little economic order whatsoever" (p. 6). He goes on to say that:

[T]o a pre-FFJR [Fama, Fisher, Jensen & Roll (1969)] observer of stock markets, the existence of systematic market behavior in response to information was almost unthinkable (Ball, 1994, pp. 31–32).

Prior to FFJR, we looked at the market reaction to information from the limited perspective of chronological time. Reading the daily financial press, for example, we observe the market reaction at a single point in chronological time to a wide cross-section of events (such as earnings and dividend announcements, ...GDP data, and rumors). We tend to see a bewildering range of events and market reactions, all at once, without seeing the correlation. FFJR inverted that perspective: they investigated the market reaction to a single event at a cross-section of chronological times. ... Where we previously had seen only

chaos, FFJR saw order. FFJR were able to report visual plots of systematic, seemingly-rational market behavior around the time of an announcement, which no previous researcher had been successful in doing (Ball, 1994, p. 31)

⁹ There are several possible explanations for the dearth of citations to Ashley. First, Ashley (1962) is a short (four-page) paper published in the *Journal of Political Economy*. Even though *JPE* was edited at the University of Chicago — the birthplace of the price-earnings research paradigm — accountants may not have been aware of the short paper that was published by a professor at the University of Cincinnati before the revolution in empirical accounting research began. Second, Ashley's methods were primitive in relation to the methodological advances in place by 1968, so accounting researchers may not have considered the study of sufficient quality to merit citation. Third, some researchers may have been reluctant to cite Ashley out of fear of diminishing the apparently novel contribution of the new price-earnings research in accounting. [Brown (1996, p. 730) highlights other instances of crediting accounting researchers for ideas developed outside of accounting.]

Benston (1967) was the first published large-scale empirical accounting study. An ambitious and complex study, especially for a pioneering effort, Benston reported a modest but statistically significant relation between ‘unexpected earnings’ and announcement month stock returns. Specifically, he concluded that a 100% change in the rate of change in unexpected earnings was associated with at best a 2% change in the monthly rate of change in security price, on average. Even though Benston was the first published accounting study to document a relation between earnings and security prices, analysis of the Social Sciences Citation Index reveals that his study did not have the impact on the literature as did the subsequent Ball and Brown (1968) and Beaver (1968) studies. Apparently, convincing the skeptics required more striking evidence than Benston was able to report given the data limitations at the time his study was conducted.¹⁰

Ball and Brown (1968) tested for an association between earnings and prices, and they found a relation between the direction of the changes in earnings and stock prices. However, they concluded that most of the information in accounting earnings was already impounded in prices by the time earnings were finally announced. Thus, while Ball and Brown’s evidence suggested that the direction of the change in earnings was value-relevant, the question of whether accounting earnings *announcements* were timely enough to convey material new value-relevant information to the market was still very much an open question.

Convincing the skeptics that at least some earnings announcements convey ‘new news’ required strong evidence.¹¹ Using a sample of non-12/31

firms and different measures of market reaction, Beaver reported squared unexpected price movements that were on average 67% higher in announcement weeks than in non-announcement weeks, and he found that trading volume increased on average by 33% during earnings announcement weeks. Increases of this magnitude are clearly ‘material’ and economically significant, and provided persuasive evidence that earnings announcements convey new information to the market.¹²

Beaver’s (1968) study, as well as Ball and Brown (1968) and Benston (1967), provided the “notable scientific achievement” that was sufficiently novel and convincing to “attract most of the next generation’s practitioners” (Kuhn, 1970, p. 18). This fundamental shift in research taste rapidly devalued the human capital of the prior generation’s normative researchers whose training preceded the empirical revolution.^{13,14} The new research paradigm offered a more attractive option for ambitious researchers intent on building a record to support promotion and tenure. In contrast to the cacophonous voices engaged in normative research, the new paradigm “provided clear-cut research

¹² Beaver (1968) made another important contribution to the development of accounting research by examining volume as well as price reactions. Perhaps because neither Benston (1967) nor Ball and Brown (1968) reported much evidence of a relation between earnings *announcements* and announcement period returns, Beaver’s research design allowed for the possibility that earnings announcements might affect either trading volume, or abnormal returns, or both. Beaver’s trading volume evidence was a necessary precursor to the later development of market microstructure studies, and his evidence that information disclosures affect prices and trading also helped spawn subsequent research modeling financial markets and their reactions to news.

¹³ We thank an anonymous reviewer whose insights provided the basis for this discussion.

¹⁴ Devaluation of the prior generation typifies the birth of a new paradigm:

[When] an individual or group first produces a [notable scientific achievement] able to attract most of the next generation’s practitioners, the older schools... disappear... [T]here are always some men who cling to one or another of the older views, and they are simply read out of the profession, which thereafter ignores their work (Kuhn, 1970, pp. 18–19)

¹⁰ Burgstahler (1987) points out that if researchers are Bayesians, their deeply ingrained prior beliefs can only be altered by powerful counter-evidence. However, researchers have alleged that Benston’s methodology “lacked power and failed to provide definitive results” (Brown, 1996, p. 730). For example, Benston related unexpected accounting numbers to *monthly* returns.

¹¹ Philip Brown (1989, p. 205) provides anecdotal evidence consistent with this view when he suggests that Ball and Brown (1968) had such an impact on subsequent accounting research because “it expressed a view that ran counter to the critics of accounting, in that it rejected the null hypothesis in a convincing way”.

problems and examples of acceptable research methods”, which gave younger researchers confidence that work within this paradigmatic framework would find acceptance (Mouck 1993, p. 41). Because the new empirical research required a set of skills that normative writers did not possess and for the most part could not substantively critique, the new paradigm offered protection in the form of barriers to entry that prevented the paradigm from being flooded with competitors and critics. Prestigious outlets such as the *Journal of Accounting Research* and its associated conference that fostered the new empirical research were effectively closed to the normative writers. The new paradigm gained cachet (largely at the expense of the prior generation of normative writers) from association with the exciting new capital markets paradigm in finance led by high-profile scholars such as Fama, Jensen and Scholes.

Kuhn (1970, p. 19) suggests that “The new paradigm implies a new and more rigid definition of the field.” The core beliefs of the emerging empirical capital markets research paradigm in accounting centered around theoretical developments in finance (e.g. portfolio theory, the capital asset pricing model, the efficient markets hypothesis), the development of large-scale event-study methodology, and the shared belief that earnings announcements were useful in making investment decisions. Researchers participating in the new capital markets paradigm soon fit Kuhn’s (p. 177) characterization of a scientific community:

To an extent unparalleled in most other fields, [members of a scientific community] have undergone similar educations and professional initiations; in the process they have absorbed the same technical literature and drawn many of the same lessons from it. Usually the boundaries of that standard literature mark the limits of a scientific subject matter.... [T]he members of a scientific community see themselves and are seen by others as the men uniquely responsible for the pursuit of a set of shared goals, including the training of their successors. Within such groups communication is relatively full and professional judgement relatively unanimous.

Once the price-earnings research paradigm was established, “a more efficient mode of scientific practice [could begin]... generally esoteric and oriented to puzzle-solving, as the work of a group can be only when its members take the foundations of their field for granted” (Kuhn, 1970, p. 178). Kuhn terms this “normal science.”

2.1.3. ‘Normal science’ and the generalization of early evidence in the price-earnings paradigm

Kuhn (1970) suggests that pre-paradigm authors begin anew developing foundations for their arguments. Such papers stand alone and so are reasonably understandable by intelligent readers. In contrast, normal science research questions are usually refinements of the pillars of the paradigm’s foundations, becoming evermore esoteric as the paradigm matures. The results of research in normal science “usually appear as brief articles addressed only to professional colleagues, the men whose knowledge of a shared paradigm can be assumed and who prove to be the only ones able to read the papers addressed to them” (p. 20).¹⁵

In normal science, the paradigm’s adherents anticipate the appropriate research methods and even the *results* of future research.¹⁶ Kaplan’s (1978, p. 168) review of information content research reveals that Kuhn’s characterization of normal science fit quite well the state of accounting research:

Current research is involved with cleaning up some of the past studies by controlling for or testing a number of factors that were not included in the original research.... At best though, these studies will confirm the findings

¹⁵ Kuhn’s observations suggest that the declining readability of accounting research may be a natural consequence of a maturing discipline.

¹⁶ Beaver (1968) identified several issues that were addressed in research spanning the next 20 years, and in each case the results were as Beaver predicted: (1) the association between firm size and market reactions to earnings announcements, (2) the relation between the magnitude of unexpected earnings and the magnitude of price reactions, and (3) the relation between the direction of the earnings news and the timing of earnings announcements (i.e. whether good news is announced earlier than bad news).

of the earlier studies by showing that the results follow even under more careful testing or using a more sophisticated methodology, and the conclusions from such studies will, therefore, be consistent with our current beliefs.

Kuhn (1970) suggests that another characteristic of normal science is the tendency to ignore inconvenient evidence, or anomalies:

And the project whose outcome does not fall in that 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. (Kuhn 1970, p. 35).

Thus, normal science is primarily engaged in solving puzzles for which the paradigm's adherents pre-suppose 'the answer'. What is at issue is not the result, but merely the ingenious means necessary to obtain the pre-supposed outcome. Kuhn's description of normal science aptly fits the academic accounting community, as shown by the following citation evidence revealing that we have collectively dismissed evidence at odds with the paradigmatic thinking in price-earnings research.

2.2. Citation analysis

The price-earnings research paradigm overlooked evidence that the information content of earnings announcements may be smaller in magnitude and less pervasive than was apparent from Beaver's original work. The literature virtually ignored the first post-Beaver published study reporting little or no market reaction to earnings announcements. Oppong's (1980) study of market reactions to 12/31 fiscal year-end NYSE firms' earnings announcements should have been of interest as a complement to Beaver's (1968) analysis of non-12/31 NYSE firms' earnings announcements. However, Oppong reports either no reaction or at best a weak reaction to 12/31 NYSE firms' annual earnings announcements. Even though Oppong's results should have been noteworthy when the study was published, and even

though the study was published as a main article in *Journal of Accounting Research*, subsequent price-earnings research has largely ignored its evidence. Table 1 shows that Oppong has received a total of only five citations in the Social Sciences Citation Index (SSCI). Panel A of Table 1 reports that of these five citations, only two refer to the study's conclusion that its sample of earnings announcements were at best only marginally informative.

The literature also largely ignored similar inconvenient evidence reported in Atiase (1985). Research typically interprets Atiase as suggesting that small firms' earnings announcements are more informative than large firms' earnings announcements. While true, this ignores the important but inconvenient fact that Atiase finds *no* significant price reaction to his large firms' earnings announcements. As Panel B of Table 1 shows, of the 27 post-1990 SSCI references to Atiase's "size effect," *none* mention that he found no reaction to large firms' announcements. While three of the pre-1990 citations acknowledge Atiase's anomalous evidence, two of these are the same papers that also acknowledge Oppong's anomalous evidence. Thus, the SSCI reveals that only three different articles (the most recent of which was published in 1989) have recognized either Oppong's (1980) or Atiase's (1985) evidence questioning the magnitude and pervasiveness of the market's reaction to earnings announcements.

Post-Beaver research has also largely overlooked two aspects of Beaver's (1968) research design that increased the chances of reporting large market reactions to accounting earnings announcements: (1) the focus on the cross-sectional mean market reaction, and (2) the sample selection criteria. To measure price reaction, Beaver (1968) employs cross-sectional means of standardized squared market model prediction errors. Extreme values can have a large impact on squared market model prediction errors, so a small number of extreme values could greatly increase the cross-sectional mean of the squared prediction errors. Citation analysis reveals that virtually none of the research citing Beaver's information content conclusion questioned whether his cross-sectional mean results were 'typical' of the reaction to individual earnings announcements. More generally,

Table 1

The tendency to ignore inconvenient evidence: analysis of citations of Oppong (1980) and Atiase (1985)^a

Panel A: Citations of Oppong (1980)					
Citation year	Acknowledges evidence of marginal reaction		Does not acknowledge evidence of marginal reaction		Total
	Research article	Review article	Research article	Review article	
1980–1990	1	1	3	0	5
1991–1996	0	0	0	0	0
Total	1	1	3	0	5

Panel B: Citations of Atiase (1985)					
Citation year	Acknowledges absence of significant reaction to large firms' earnings announcements		Does not acknowledge absence of significant reaction to large firms' earnings announcements		Total
	Research article	Review article	Research article	Review article	
1985–1990	2	1	27	2	32
1991–1996	0	0	24	3	27
Total	2	1	51	5	59

^a Citations of Oppong (1980) and Atiase (1985) listed in the Social Sciences Citation Index. To be included in this analysis, the citations must cite the studies in an information content context.

despite Ball's (1994, p. 28) query: "How frequently is a sample average residual dominated by a small number of outliers...?" subsequent research has not addressed the *pervasiveness* of the market's reaction to earnings announcements.

Beaver employed sample selection criteria that 30 years of hindsight suggest effectively increased the chances of finding significant market reactions to earnings announcements. His sample included earnings announcements of non-12/31 NYSE firms with fewer than 20 news items in the *WSJ* during the year. Beaver (1968, pp. 70–71) limited the analysis to non-12/31 firms to avoid event clustering,¹⁷ and invoked the 20 news item limit to

obtain a cleaner non-announcement period. Non-12/31 firms are on average smaller than 12/31 firms (Smith & Pourciau, 1988), and Atiase (1985) and Bamber (1987) show that price and volume reactions to small firms' earnings announcements exceed the reactions to larger firms' announcements. Also, Grant (1980) and Shores (1990) provide evidence that the magnitude of the price reaction to earnings announcements is a decreasing function of the number of *WSJ* announcements. Thus, Beaver's analysis was confined to smaller firms with fewer news items (i.e. less pre-disclosure information), which in turn increased the likelihood of a significant reaction to earnings announcements.¹⁸

In the historical context in which the research was conducted, this directed sampling approach (which was likely unintentional at the time) was

¹⁷ Most researchers suggest that they use this criterion to avoid underestimation of the standard deviation in the presence of cross-sectional correlation. On the other hand, Beaver's rationale is that "When earnings announcements cluster, ... any attempt to remove the effects of market-wide events would eliminate [some of] the effects of the earnings report as well" (p. 71). However, even if clustered, earnings announcements should not affect market returns if positive returns roughly offset negative returns.

¹⁸ The next information content study, May (1971), is based on a sample of 105 AMEX firms. Thus, the first two information content studies used firms that were on average smaller than the typical NYSE firm, and as a result, these two studies' earnings announcements were more likely to be informative.

reasonable. Wallace (1991, p. 35) suggests that “directed sampling” can be appropriate when “researchers will design the study to maximize the ‘affected’ group... on the premise that, if an effect is not discernable for that group, then it is unlikely it would be ... of importance in drawing inferences to a wider group.” However, she cautions that “The price of directing the study is that the inferences drawn must be similarly directed to the population from which the sample was selected.”

While Beaver’s evidence was sufficient to refute the view that earnings announcements were useless to investors, the magnitude of the reported market reaction was unlikely to be representative of a broader population. It is important to note, however, that *Beaver carefully pointed out the likely bias induced by at least one of the sample selection criteria, and he cautioned against generalizing the study’s results to other populations of earnings announcements:*

It is possible that the selection criteria, especially criterion (6) [that there be less than 20 news announcements per year in the *WSJ*] may induce some bias in the [information content] direction. As long as the criteria are visible ex ante, the population for which the study’s findings are relevant can be easily identified (p. 72).

Nevertheless, *most subsequent researchers’* references to Beaver’s results did generalize, implicitly or explicitly, without acknowledging the likely threats to external validity imposed by the sample selection criteria. This is consistent with Kuhn’s (1970, p. 178) observation that adherents of normal science paradigms “take the foundations of their field for granted” in subsequent research.

Table 2 tabulates the references to Beaver’s (1968) conclusion that his sample earnings announcements are informative, on average. Most of these references (123) were identified from the Social Sciences Citation Index (SSCI), and 9 were identified from other sources including books, monographs, and research reports.¹⁹

Panel A categorizes the citations by source (SSCI or Other), by type of citing source (research article or review), and by the manner in which Beaver (1968) is cited. Other than Davidson’s (1968) discussion of Beaver’s paper at the *Journal of Accounting Research* conference, these citations did not question whether the cross-sectional mean results were ‘typical’ of the reaction to individual earnings announcements. Consequently, Table 2 partitions based on acknowledgment of potential effects of the second research design choice — the sample selection criteria. Citations that implicitly or explicitly acknowledge the potential effects of Beaver’s sample selection criteria are listed on the left side of the table, while citations that fail to do so are shown on the right. Panel A further partitions citations failing to acknowledge Beaver’s sample selection criteria into those that refer to the specific numerical magnitude of his observed reactions, and/or reprint one or more of the well-known figures (e.g. Beaver’s Figs. 1, 3 or 6), versus those that cite Beaver’s results in a general, qualitative fashion.

Panel A reveals that only 8 of the 132 total citations (6.1%) acknowledge the potential effects of Beaver’s sample selection criteria, and more important, only 2 of the 29 (2+8+19) reviews (6.9%) do so. The low proportion of reviews that acknowledge Beaver’s sample selection criteria is disturbing because reviews distill ‘stylized facts’ — misleading or incomplete statements are likely to stimulate unwarranted generalization. Eight of the

¹⁹ We found 245 citations of Beaver (1968) in Social Sciences Citation Indexes. Of these, 6 were not written in English, 20 did not actually refer to Beaver (1968) in the text even though they listed Beaver (1968) among the references, and 96 referenced Beaver (1968) for reasons unrelated to the conclusion that earnings announcements are informative. Table 2 categorizes the remaining 123 citations, plus nine additional citations from other sources, including: Dyckman, Downes and Magee (1975), Foster (1986), Dyckman and Morse (1986), Watts and Zimmerman (1986), Griffin (1987), Bernard (1989), Beaver (1989), Bernard

(1991), and Ball (1994). More detailed analysis of these 132 citations to Beaver’s (1968) information content result reveals that the paper has been cited in a substantive fashion by a wide cross-section of citing authors. For example, 170 different authors were involved in the 132 citing sources, and only 14 (10%) of the citing articles were authored by Beaver or a Stanford Ph.D. Moreover, our (necessarily subjective) evaluation of each citation revealed that 130 of the 132 citations of Beaver’s information content inference are in our opinion clearly substantive, warranted citations, given the context of the citing research.

Table 2
Analysis of citations of Beaver's (1968) information content inference

Panel A: By source and article type							
Citation source	Acknowledges effects of Beaver's sample selection criteria ^c		Does not acknowledge effects of Beaver's sample selection criteria				Total
	Research article	Review article	Reproduces Beaver's metric values or graphs ^d	Does not reproduce Beaver's metric values or graphs	Research article	Review article	
SSCI ^a	6	0	7	3	90	17	123
Other ^b	0	2	0	5	0	2	9
Total	6	2	7	8	90	19	132

Panel B: By year and article type					
Citation year	Acknowledges effects of Beaver's sample selection criteria ^c		Does not acknowledge effects of Beaver's sample selection criteria		Total
	Research article	Review article	Research article	Review article	
1970–1974	0	0	11	9	20
1975–1979	0	0	20	3	23
1980–1984	2	0	24	3	29
1985–1990	3	2	21	7	33
1991–1996	1	0	21	5	27
Total	6	2	97	27	132

^a English language citations of Beaver (1968) listed in the Social Sciences Citation Index. To be included, the citation must reference Beaver's conclusion that earnings announcements are "informative".

^b 'Other' citations located from various sources. These include Dyckman et al., (1975), Foster (1986), Dyckman and Morse (1986), Watts and Zimmerman (1986), Griffin (1987), Bernard (1989), Beaver (1989), Bernard (1991), and Ball (1994).

^c Citations that do not generalise Beaver's (1968) results without implicitly or explicitly acknowledging the potential effects of Beaver's (1968) sample selection criteria.

^d Citations that reproduce Beaver's (1968) announcement week price and/or volume metric values, or else reproduce his price and/or volume graphs.

29 reviews (27.6%) refer to the specific magnitude of Beaver's reported average reactions, a finding that is unlikely to arise in other samples. Six reviews (20.7%) reprinted one or more of Beaver's figures, where the large earnings announcement week spikes provide salient visual cues suggesting that earnings announcements are associated with strong market reactions. Evidence reported in Panel B reveals that the tone of these references has not changed in recent years. Beaver (1968) continues to be heavily cited, and 96% (26÷27) of the references in the most recent six-year period fail to acknowledge the likely effects of Beaver's sample selection criteria. Thus, the received wisdom in the price-earnings paradigm is that the market reacts strongly to accounting earnings

announcements.²⁰ In sum, our citation analysis shows that post-Beaver normal science has largely overlooked both (1) counter-evidence reported in Oppong (1980) and Atiase (1985), and (2) aspects

²⁰ For example,

Beaver's paper is highly cited because it showed that earnings are informative for both stock price variance and trading volume (Brown 1996, p. 730)

It has long been documented that there is much unusual stock price activity immediately surrounding earnings announcements (Bernard 1991, p. 305).

One of the most robust findings in the financial statement research area is that the release of interim and annual earnings is associated with both increased trading volume and increased security return variability (Foster, 1986, p. 377).

of Beaver's research design (focus on the cross-sectional mean reaction and sample selection criteria) that likely increase the magnitude of the observed market reaction to earnings announcements.

We investigate effects of two alternative research design choices that would have been reasonable at the time Beaver (1968) was conducted. First, rather than simply reporting the cross-sectional mean market reactions, we also investigate the pervasiveness of market reactions to earnings announcements. Specifically, we provide new evidence on the proportion of earnings announcements that stimulate an unusual market reaction. This statistic is not unduly influenced by a few extreme observations, in contrast to the cross-sectional mean statistic. This analysis will help determine whether: (1) only a few earnings announcements generate large price reactions while most stimulate little unusual price activity, or (2) most earnings announcements stimulate a modest amount of unusual price activity. Second, during the mid-1960's when there was no empirical evidence on the information content of *any* firms' announcements, there would have been interest in the informativeness of major firms' earnings announcements. Therefore, in addition to examining reactions to earnings announcements made by firms meeting Beaver's sample selection criteria, we also investigate reactions to earnings announcements made by the era's Fortune 200.

3. Research method

3.1. Sample selection

We first develop a set of firms comparable to Beaver (1968). Beaver analyzed reactions to 506 annual earnings announcements from 1961 to 1965, made by 143 non-12/31 NYSE firms listed on COMPUSTAT. He eliminated earnings announcements made by firms about which there were 20 or

more news items in the *WSJ* during the year,²¹ and he also eliminated announcements that were potentially confounded by dividend announcements during the earnings announcement week or stock splits announced within 17 weeks centered on the earnings announcement week. These criteria led to a sample of smaller-than-average NYSE firms that constituted only 5.1% of the NYSE market value.²²

Approximating Beaver's set of firms 30 years later requires considerable care. The Appendix explains the details of our sample construction procedure. We identified an initial group of 255 non-12/31 NYSE firms that had CRSP data and were listed on COMPUSTAT during 1962–1966. This is very close to Beaver's initial set of 242 non-12/31 NYSE firms with COMPUSTAT and CRSP data, as reported in his Table 1. For these 255 firms, we found a total of 1,079 earnings announcements in the *WSJ*, over fiscal years 1962–1966. Hereafter, we term these the 'Beaver-like announcements'.²³

For the alternative sample, we identified the Fortune 200 firms during 1962–1966.²⁴ The *WSJ* listed a total of 941 earnings announcements for these firms, for fiscal years 1962 to 1966. Given the paucity of research on the information content of *any* firm's earnings announcements at the time

²² For Beaver's sample we calculate: $(143 \text{ firms}) \times (\$189 \text{ million average market value of a sample firm's common shares outstanding, per p. 71 of Beaver}) \div (\$528,190 \text{ million total market value of NYSE common shares outstanding}) = 5.1\%$ of NYSE market value.

²³ Our Beaver-like sample of 1,079 annual earnings announcements by non-12/31 firms is a very different sample from Atiase's (1985) "small NYSE/AMEX firm" sample, which includes 50 second quarter earnings announcements of the very smallest 12/31 fiscal year-end firms (less than \$20 million market value). The average market value of our Beaver-like firms is over 7 times higher than the \$20 million upper limit on Atiase's 'small firm' sample.

²⁴ Twenty-three of the Fortune 200 firms also meet the criteria for our Beaver-like firms. Since our purpose is to illustrate the effects of different sample selection criteria (rather than to partition one sample into mutually exclusive sub-samples), we include the 23 firms in both the Beaver and the Fortune 200 groups. This works against finding significant differences between the two groups. One of the Fortune 200 firms, Anheuser-Busch, was traded on the Nasdaq until 1980, so this firm's earnings announcements are omitted from the analysis.

²¹ In his random sample of 100 NYSE firms in the period 1960–1964, Grant (1980) reports that the median number of *Wall Street Journal* news items is 14. Beaver's 20 item cutoff is well above the typical number of announcements per firm during that period.

Beaver (1968) was conducted, evidence on the Fortune 200 would have been of interest, as these firms constituted two-thirds of the market value of the NYSE during that time.²⁵

Our analysis follows Beaver's approach, with two minor exceptions. First, while Beaver's data covered 1961 to 1965, we use data from 1962–1966 because this is the earliest period for which CRSP data are currently available. Second, Beaver eliminated earnings announcements made by firms about which there were 20 or more *WSJ* news items during the year. This criterion eliminated less than 20% of Beaver's data, since his sample included only non-12/31 firms which are on average smaller and less intensely covered by news media.²⁶ However, application of this news restriction to the intensely covered Fortune 200 would not only eliminate most of the earnings announcements (nearly 60%), but it would eliminate those announcements for which the most predislosure information was available, leading to a biased sample of earnings announcements that are more likely to be informative. Thus, we do not apply Beaver's 20 news item criterion in the analysis reported formally in our tables, although our discussion will explain the results we obtain upon applying this criterion. The analysis reported here does apply Beaver's stock split and dividend announcement restrictions, which reduce the Beaver-like sample from 1,079 to 854 announcements, and the Fortune 200 sample from 941 to 664 announcements. (Analysis of the full samples of 1,079 and 941 announcements yield results that are very similar to those reported here.)

Beaver presented evidence on both price and volume reactions to earnings announcements. While we analyzed both price and volume reactions, the inferences from volume reactions are

quite similar to those based on price reactions. Hence, we confine the discussion to price reactions.

3.2. Announcement and parameter estimation periods

Earnings announcement dates are obtained from the *Wall Street Journal Index*. We examine weekly price reactions over a 17 week period centered on the earnings announcement week.²⁷ The parameter estimation period includes all weeks other than the 17 week announcement periods, as in Beaver (1968).

3.3. Price reaction metric

The price reaction metric employs prediction errors from a firm-specific market-model:

$$u_{iw} = R_{iw} - [\hat{a}_i + \hat{b}_i R_{mw}]$$

where:

- u_{iw} = firm *i*'s unexpected week *w* return,
- R_{iw} = natural log of firm *i*'s week *w* return,
- R_{mw} = natural log of the return on Standard and Poors' Index (as in Beaver 1968) during week *w*, and
- \hat{a}_i, \hat{b}_i = firm *i*'s market model parameter estimates, estimated over the parameter estimation period.

Beaver squared the unexpected return (to abstract from its sign), and normalized the resulting price

²⁵ Our sample of 941 annual earnings announcements of Fortune 200 firms differs from Atiase's (1985) 'large firm' sample, which is based on 100 second quarter earnings announcements of 12/31 fiscal year-end 'large firms'. Researchers may have discounted Atiase's evidence of virtually no reaction to large firms' earnings announcements because of concerns about the power of analyses based on only 100 observations.

²⁶ Application of the 20 news item restriction reduces our Beaver-like sample by about 15%.

²⁷ Announcements appearing in the *WSJ* on Mondays are typically announced on the Broad Tape on the preceeding Friday. Hence, for Monday announcements, the week prior to the *WSJ* announcement is the earnings announcement week.

²⁸ While squaring exacerbates the effect of extreme values, the problem identified in our study is not simply attributable to this effect of the squaring operation. We obtain qualitatively similar inferences using May's (1971) price reaction metric, which is less sensitive to extreme values in both the announcement and estimation periods, because it is defined as the *absolute value* of the announcement period unexpected return divided by the (firm-specific) mean of the *absolute values* of the estimation period unexpected returns.

reaction measure by the variance of estimation period unexpected returns.²⁸ He labeled this the U ratio:

$$U = \frac{(u_{iw})^2}{(s_i)^2}$$

and s_i^2 is the estimated variance of firm i 's weekly unexpected returns, calculated over the parameter estimation period. Following Beaver (1968), we report cross-sectional mean U statistics.

Beaver's U ratio measures announcement week price movement relative to the average price movement in the estimation period. If the ratio exceeds one, the announcement period price movement is higher than normal, and vice versa.²⁹ Beaver (1968) and subsequent researchers have considered 1.0 as a benchmark hurdle for inferring that U's (price movements) are abnormally high.

4. Results

4.1. Beaver-like sample

The left-hand side of Table 3 reports cross-sectional mean U's for our Beaver-like firms' earnings announcements throughout the 17 week period centered on the earnings announcement date. Fig. 1 plots these results. There is a strong spike at week 0, and this announcement week U is significantly higher ($p \leq 0.05$) than the U in each of the surrounding weeks (except for week +5). The

announcement week cross-sectional mean U is 1.453 — about 45% higher than its expected value.³⁰ Given Beaver's (1968, p. 79) benchmark that U's greater than 1.0 mean that the "price change is larger than normal," our results confirm Beaver's original conclusion that, *on average*, earnings announcements for non-12/31 NYSE firms are associated with unusual price movements. However, the cross-sectional average does not answer other important questions. How pervasive are market reactions of this magnitude? What proportion of earnings announcements convey new information to the market and cause any unusual announcement week price reaction?

To address these questions, the third column of Table 3 presents the proportion of Beaver-like earnings announcements whose announcement week price variability exceeds the firm-specific average non-announcement price variability (i.e., where U's exceed 1.0).³¹ This reveals a very different picture than the cross-sectional mean U's. Only 33.7% of the earnings announcements generate U's exceeding 1.0, which is not very impressive given that on average 25.8% of the U's exceed 1.0 even in the surrounding 16-week

²⁹ Patell (1976) shows that the U's theoretical expected value is slightly greater than 1.0 (1.02 in our study) due to prediction error and properties of the F distribution.

³⁰ The 20 news item restriction does not affect inferences for the Beaver-like firms, since the announcement week cross-sectional mean U is 1.4 even after applying this restriction. Thus, our announcement week cross-sectional mean U is slightly lower than the 1.67 reported by Beaver (1968). Despite efforts to ensure that our firms are as close as possible to Beaver's (see our Appendix), there may be some differences. Also, Beaver's time period spanned 1961–1965, while our sample includes earnings announcements made after July 1962 (the beginning of the CRSP tapes) up to announcements of fiscal years ending by December 1966.

³¹ Little is known about the true empirical time-series or cross-sectional distributions of U ratios. Patell (1976) showed that the within-firm time-series distribution of a firm's U ratios is a scaled F distribution (scaled to account for predicting announcement period unexpected returns outside the parameter estimation period), provided that the market model residuals are normally distributed. Under this assumption, the theoretical expectation is that, even in the absence of an announcement, 32% of a given firm's weekly non-announcement period U's would exceed 1.0. However, Marais (1984) pointed out that the (within-firm) time-series distribution of returns is leptokurtic for many firms. This affects the variance, but not the mean of the U distribution. Marais showed that for normally distributed returns the asymptotic variance of the U ratio is 2, whereas for leptokurtic returns the asymptotic variance of U equals $2 + \text{the kurtosis of estimation period market model residuals}$. Using Patell's earnings forecast data and a random sample of 1000 weekly returns for NYSE firms, Marais found that the kurtosis of these residuals ranged between 2.5 and 4.0. Thus, dispersion among U ratios was higher than implied by Patell's (1976) analysis, indicating that there are apt to be more of the highly extreme values in the upper tail of the distribution of U ratios, and more values close to zero, than occurs in a scaled F distribution.

Table 3

Price reactions to earnings announcements of Beaver-like firms and Fortune 200 firms

Week relative to earnings announcement	Beaver-like earnings announcements ^a		Fortune 200 earnings announcements ^c	
	Cross-sectional mean U-statistic	% Earnings announcements with U-statistic greater than 1	Cross-sectional mean U-statistic	% Earnings announcements with U-statistic greater than 1
–8	0.841	21.7%	1.076 ^d	29.8% ^d
–7	0.944	23.2%	1.143 ^d	27.7% ^d
–6	0.974	25.1%	0.972 ^d	26.4% ^d
–5	0.804	25.6%	1.038 ^d	29.2% ^d
–4	0.943	22.7%	1.004 ^d	28.6% ^d
–3	0.882	23.5%	0.895	25.3%
–2	0.903	24.5%	0.909	27.9% ^d
–1	0.929	25.5%	1.002 ^d	28.0% ^d
0	1.453	33.7%	1.179	30.0%
+1	1.204	29.0%	1.006 ^d	28.8% ^d
+2	1.084	24.9%	0.954	25.5%
+3	0.977	26.6%	0.963	24.5%
+4	1.095	27.9%	0.960	27.2% ^d
+5	1.232 ^b	28.5%	1.020 ^d	25.8%
+6	1.074	27.6%	1.064 ^d	27.6% ^d
+7	1.043	29.4%	1.000 ^d	28.5% ^d
+8	1.027	27.3%	0.962 ^d	26.2% ^d
Average week excluding week 0	0.997	25.8%	0.998	27.3%

^a Earnings announcements for fiscal years 1962 to 1966 made by our ‘Beaver-like’ non-12/31 NYSE firms. Following Beaver (1968), we exclude earnings announcements that are confounded by a dividend announcement in the earnings announcement week, or a stock split in the 17 weeks centred on the earnings announcement week.

^b This week’s U-statistic and the announcement week’s U-statistic are *not* significantly different at $p < 0.05$, by the matched-pair *t*-test.

^c Fortune 200 firms’ earnings announcements for fiscal years 1962–1966, provided that the announcement is *not* potentially confounded by a dividend announcement in the earnings announcement week, or a stock split in the 17 weeks centred on the earnings announcement week.

^d This week’s statistic and the announcement week’s statistic are *not* significantly different from zero at $p \leq 0.05$, by the matched-pair *t* test (for U-statistics), or by the confidence interval about a sample proportion (for % earnings announcements with $U > 1$).

non-announcement period.^{32,33} Thus, even for the Beaver-like announcements, significant cross-sectional market reactions at the announcement date are driven by the specific reactions to a relatively small number of announcements. Focus on the *cross-sectional average* market reaction has obscured the fact that even for the Beaver-like

firms, *most* of the earnings announcements generate little unusual price variability.

However, requiring U’s to exceed 1.0 is a low, or easily achievable benchmark. Alternative benchmarks for assessing unusual price movements might consider U’s unusually high only if they are in the top 5–10% of the distribution of firm-specific non-announcement period U’s. Relative to

³² Beaver (1968) focusses on cross-sectional mean results, and does not report the percentage of U’s exceeding 1.0. However, the study does report sufficient data to enable the reader to make this calculation. Page 82 reveals that 181 of the announcement week U’s exceed 1.0. Combining this with the sample size of 506 announcements (page 72) reveals that $181/506 = 35.8\%$ of the earnings announcements generated U’s exceeding 1.0. The citation analysis described in Table 2 did not reveal a single reference to this implication. More generally, subsequent research did not question whether the cross-sectional

tional mean results were representative of reactions to typical individual earnings announcements. While Beaver (1968) reports sufficient data to raise a question about the pervasiveness of the market reaction to earnings announcements, our citation evidence indicates that subsequent researchers overlooked these clues.

³³ The proportion of U’s exceeding 1.0 is significantly higher in the announcement week than in each of the surrounding 16 non-announcement weeks ($p < 0.03$).

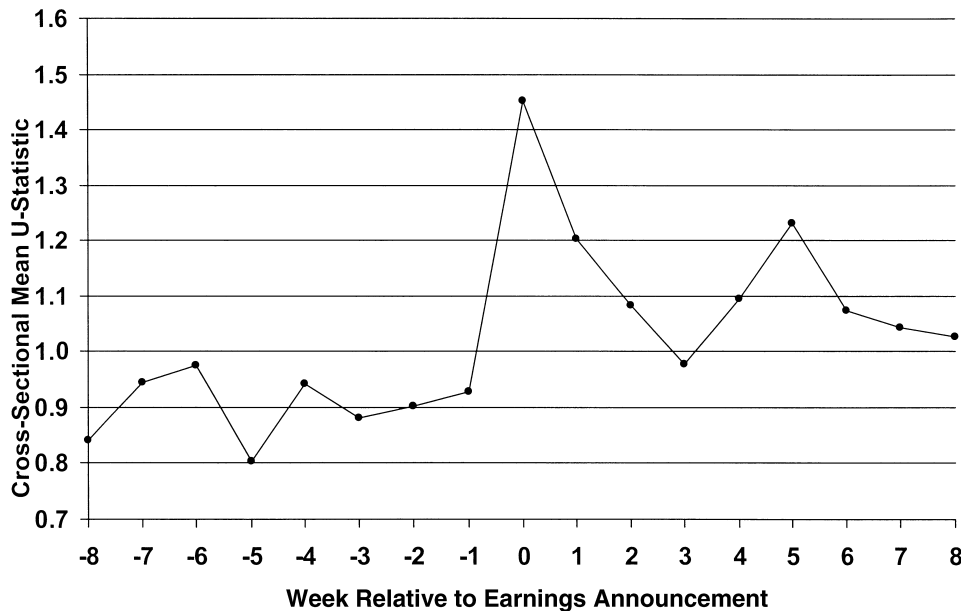


Fig. 1. Price reactions (cross-sectional mean U-statistics) to Beaver-like earnings announcements.

the 1.0 benchmark, such alternative and more demanding benchmarks classify an even smaller proportion of price reactions as abnormally high. We construct such alternative firm- and announcement-specific benchmarks by ranking a firm's announcement week U relative to the distribution of the firm's U's from the surrounding 16 non-announcement weeks.³⁴ For each earnings announcement, we performed the following steps:

1. For the 17 week period centered on the firm's earnings announcement week, rank the firm's U's from weeks -8 to +8 so that the week with the highest U is assigned rank 1 and the week with the smallest U is assigned rank 17.
2. Perform a cross-sectional analysis of the ranks of earnings announcement week U's.

2(A) Compute the proportion of sample earnings announcement week U's that are assigned rank 1, the proportion of announcement week U's assigned rank

2, and so on to the proportion of announcement U's assigned rank 17.

- 2(B) Construct a frequency histogram displaying the proportion of earnings announcement week U's assigned rank 1, the proportion assigned rank 2, and so on to the proportion assigned rank 17.

Fig. 2 depicts the percentage of earnings announcement week U's that are assigned each of the 17 ranks. Under the null hypothesis of no information content, 1/17 (about 5.9%) of the earnings announcement week U's fall in each of the 17 ranks. If earnings announcements are informative, the announcement week U's should concentrate in ranks 1 and 2. However, Fig. 2 shows that only 10.8% of the announcement week U's are assigned rank 1 (i.e. generate U's that are the highest in their 17 week period). Similarly, only 7.5% of the earnings announcement week U's are assigned rank 2. The proportion of earnings announcement week U's assigned each of the other 15 ranks is not significantly different from that expected in the absence of an announcement (except that the proportion assigned rank 14 is

³⁴ We thank George Foster for suggesting this analysis.

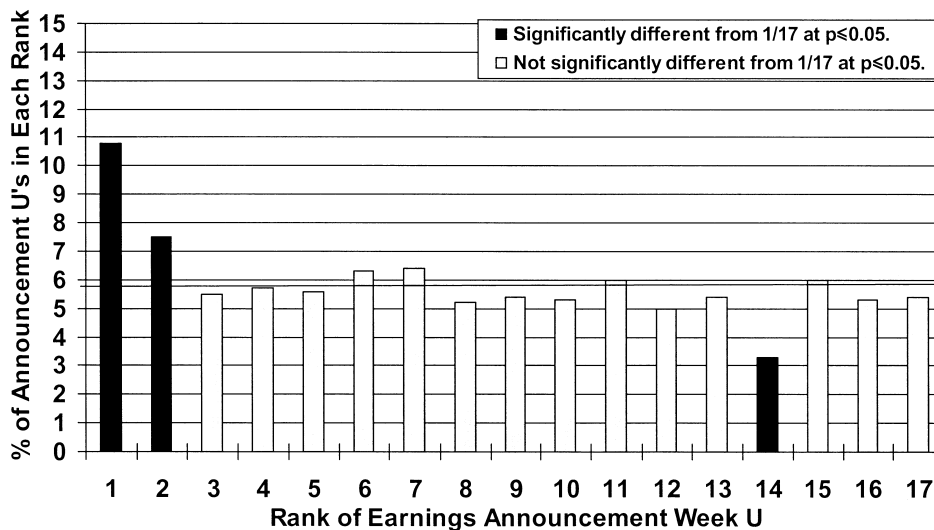


Fig. 2. Ranking of Beaver-like firms' announcement week price movements (U-statistics) relative to weekly price movements in the surrounding (nonannouncement) 16 weeks. For the 17 week period centered on the firm's earnings announcement, we rank the firm's weekly U's from weeks -8 to $+8$, so that the week with the highest U is ranked 1, and the week with lowest U is ranked 17. The figure displays the % of announcement week U's assigned rank 1, the % assigned rank 2, and so on to the % assigned rank 17. In the absence of an announcement, 5.88% ($1/17$) of the weekly U's would be expected to fall in each of the 17 ranks.

below the expected 5.9%). The histogram confirms that even for the Beaver-like announcements, the information content inference is driven by a small proportion of earnings announcements. Only 18.3% ($10.8 + 7.5\%$) of the earnings announcement U's are assigned ranks 1 and 2, which is not very impressive given that 11.8% ($2 \times 5.9\%$) of the U's would be expected in these two ranks even in the absence of an earnings announcement.

4.2. Fortune 200 sample

Evidence regarding market reactions to the Fortune 200 firms' earnings announcements appears in the right-hand side of Table 3, and in Fig. 3. The Fortune 200 cross-sectional mean announcement period U is 1.179 — about 18% higher than its expected value of 1. However, this is less than half the magnitude of the 45% reaction to the Beaver-like announcements.³⁵ Fig. 3

illustrates this difference by superimposing plots of the Beaver-like versus Fortune 200 U's. In addition to being smaller in magnitude, there is less evidence that Fortune 200 announcement week reactions are unique, since the magnitude of the cross-sectional mean announcement week U is not significantly different from 11 of the other 16 weeks' cross-sectional mean U's.

To investigate the pervasiveness of market reactions to the Fortune 200 earnings announcements, the rightmost column of Table 3 presents the proportion of U's that exceed 1.0. Only 30% of these earnings announcements generate U's exceeding 1.0, an even smaller proportion than the Beaver-like earnings announcements. This is not very impressive, given that in *non-announcement periods*, 27.3% of the weekly U's exceed 1.0. Thus, the significant cross-sectional average market reaction is driven by the specific reactions to an even smaller proportion of earnings announcements than in the Beaver-like sample. Table 3 also reveals that the proportion of U's exceeding 1.0 is generally not uniquely high in the earnings announcement week, since the announcement period proportion is not

³⁵ The announcement week U is statistically significantly higher for the Beaver-like earnings announcements than for the Fortune 200 announcements ($p < 0.02$).

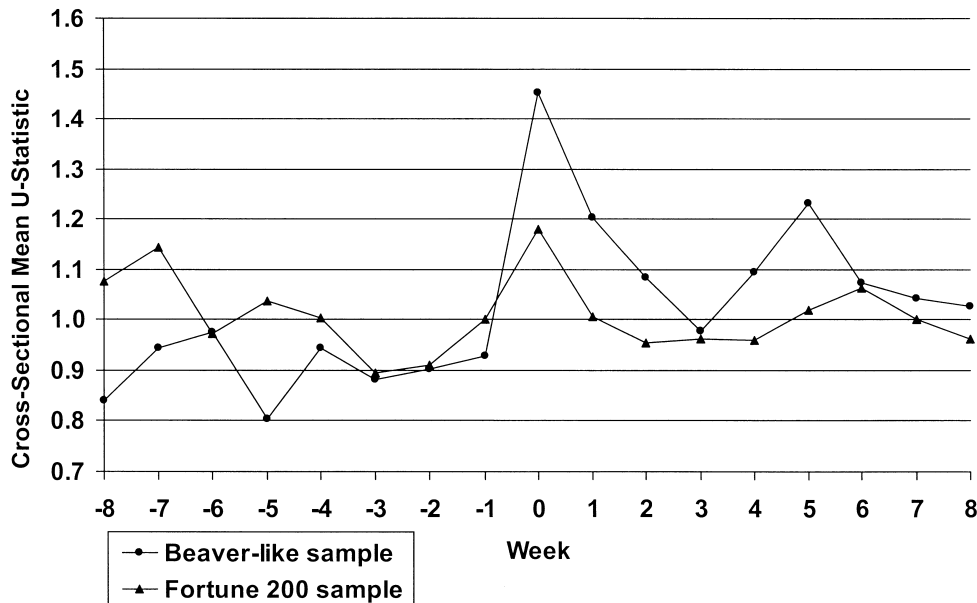


Fig. 3. Comparison of price reactions (cross-sectional mean U-statistics) between Beaver-like and Fortune 200 earnings announcements.

significantly different from 12 of the 16 surrounding non-announcement weeks.

To further assess the pervasiveness of the market reaction, for each of the Fortune 200 announcements we rank the U's from weeks -8 to +8, where the week with the highest U is ranked 1, and the week with the lowest U is ranked 17. Fig. 4 reports the proportion of Fortune 200 earnings announcement week U's in each of the 17 ranks. The evidence is even more striking than analogous evidence for the Beaver-like announcements. *For the Fortune 200, the proportion of announcement week U's assigned each of the low ranks is not significantly different from the 5.9% expected in the absence of*

an earnings announcement! Only 7.3% of the announcement week U's are assigned rank 1 (generate U's that are the highest in the 17 week period). This is not significantly different from the 5.9% expected in the absence of an announcement. Except for rank 14 where there are fewer announcements than expected under the null, the proportion of earnings announcement U's falling in each of the ranks is not significantly different from the 5.9% expected in the absence of an announcement. The histogram confirms that the information content inference regarding the Fortune 200 earnings announcements is attributable to a small proportion of the announcements.³⁶

³⁶ Results presented in Table 3 and Figs. 1–4 are based on earnings announcements that are not confounded by dividend announcements or stock splits. We did not apply Beaver's 20 news item restriction in this analysis because the news criterion is not related to contemporaneous contaminating 'events', and it eliminates firms with significant public predislosure information. Eliminating earnings announcements of firms for which the *WSJ* has published 20 or more items during the year has little effect on our results for the Beaver-like announcements, as only 15% are eliminated. In contrast, the Fortune 200 are heavily covered by the *WSJ* and the news criterion eliminates nearly 60% of our Fortune 200 announcements. For the remaining Fortune 200 firms there are fewer alternative

sources of information, so their earnings announcements are more likely to be informative. Applying the 20 news item restriction defeats the purpose of using the Fortune 200 to assess the generalizability of Beaver's results to a different and economically important population of firms. Consequently, it is not surprising that the market response to these 'neglected' Fortune 200 firms' earnings announcements is similar to the reaction to Beaver-like firms' announcements (the announcement week cross-sectional mean U for the 'neglected' 40% of the Fortune 200 announcements is 1.3, and 33% of these announcements' U's exceed 1.) Thus, the market reaction to Fortune 200 earnings announcements is sensitive to the predislosure information environment.

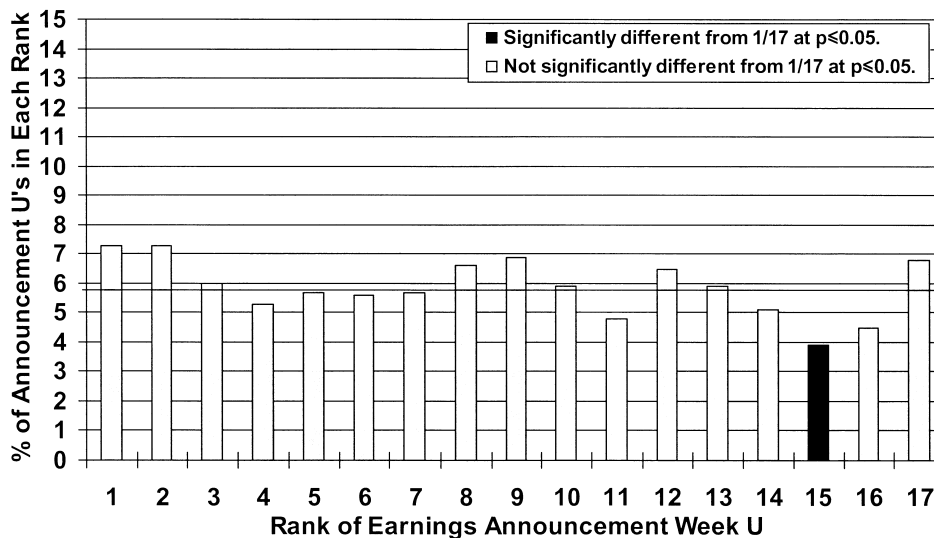


Fig. 4. Ranking of Fortune 200 firms' announcement week price movements (U-statistics) relative to weekly price movements in the surrounding (nonannouncement) 16 weeks. For the 17 week period centered on the firm's earnings announcement, we rank the firm's weekly U's from weeks -8 to $+8$, so that the week with the highest U is ranked 1, and the week with the lowest U is ranked 17. The figure displays the % of announcement week U's assigned rank 1, the % assigned rank 2, and so on to the % assigned rank 17. (In the absence of an announcement, 5.88% ($1/17$) of the weekly U's would be expected to fall in each of the 17 ranks.)

In conclusion, for both the Beaver-like and the Fortune 200 announcements, focus on the cross-sectional mean market reaction has obscured the fact that *there is little evidence of an unusual price reaction for most of the earnings announcements*. The evidence reported here, especially in Figs. 2 and 4, questions the validity of the 'stylized fact' that accounting earnings announcements typically convey new information to the market.

5. Summary, conclusions, and implications

While accounting scholars 'think we understand' that the results and generalizability of research studies are sensitive to research design choices, this paper provides evidence that we have collectively failed to apply this understanding in practice. Consistent with Kuhn's (1970) observation that a paradigm's disciples take for granted their paradigm's foundations, we find that the community of accounting researchers has failed to

appreciate fully the likely effects of the (necessarily) subjective research design choices in Beaver's (1968) seminal paper *The Information Content of Annual Earnings Announcements*.³⁷ Our citation analysis shows that subsequent research misinterpreted Beaver's evidence: (1) by viewing the cross-sectional mean reaction as 'typical' of the reaction to individual earnings announcements, and (2) by failing to appreciate the implications of the sample selection criteria. The citation analysis also reveals that subsequent research largely overlooked counter-evidence reported by

³⁷ Our analysis is *not* a criticism of Beaver's (1968) seminal study. Rather, we raise concerns about subsequent researchers' interpretations of Beaver's evidence. We could have selected studies other than Beaver (1968) to illustrate subsequent misinterpretations or overlooking of evidence, such as May (1971), Oppong (1980), or Atiase (1985). Another example of this phenomenon is the way researchers for many years virtually ignored the post-announcement drift apparent in Ball and Brown's (1968) Fig. 1.

Oppong (1980) and Atiase (1985), consistent with Kuhn's (1970) view that adherents of a paradigm ignore evidence counter to the paradigm's core beliefs. We then report results of new empirical analyses that provide evidence relevant to these misconceptions.

Our analyses question the extent to which earnings announcements convey new information to the market. With respect to the first misinterpretation of Beaver's results, we find that — for the smaller Beaver-like firms as well as for the larger Fortune 200 firms — *there is no evidence of a significant price reaction for the majority of earnings announcements*. This new evidence suggests that the focus on cross-sectional means has obscured the fact that significant mean reactions have been driven by a small proportion of announcements, and in fact, there is no evidence that most earnings announcements convey new information to the market.

With respect to the second misinterpretation of Beaver's (1968) results, the vast majority of subsequent research fails to acknowledge that Beaver's sample selection criteria (non-12/31 firms with less than 20 other news items in the *WSJ* that year) led to a sample of smaller-than-average firms. In fact, Beaver's sample constituted only about 5% of the NYSE's total market value at the time. To provide evidence on the effect of these sample selection criteria on the generalizability of Beaver's results, we assessed the information content of earnings announcements made by an alternative set of firms — the Fortune 200. At the time Beaver's (1968) study was conducted, there was no evidence on the informativeness of *any* firms' earnings announcements, so the information content of such major firms' earnings announcements would have been of interest, particularly since the Fortune 200 constituted two-thirds of the NYSE's total market value. While we find a small cross-sectional mean reaction to the Fortune 200 earnings announcements, the magnitude of this reaction is statistically indistinguishable from price movements in several of the surrounding weeks. Moreover, relative to the Beaver-like announcements, an even smaller proportion of the Fortune 200 earnings announcements generate unusual price reactions, and

depending on the benchmark, in some cases there is no evidence that the proportion of Fortune 200 firms with unusual market activity differs from that expected in the absence of an announcement.

These misinterpretations of Beaver's (1968) results are important because Beaver's evidence (in combination with May's (1971) evidence on the information content of AMEX firms' earnings announcements) formed the basis for the 'stylized fact' that earnings announcements cover new information to the market. We now consider how several facets of the practice of science in the academic accounting community combined to foster an environment where 'the placement of the first research bricks affected the whole wall'.

The first set of facets includes two cognitive biases to which individual researchers may be subject. As Hogarth (1987, p. 268) says of Kahneman and Tversky's (1972) representativeness heuristic, "We tend to imagine that what we see... is *typical* of what can occur. We seldom give credence to the possibility of 'surprising' or 'unusual' events. Hence, we tend to assess distributions which are too tight." Researchers susceptible to the representativeness heuristic are unlikely to appreciate that the mean effect may not be representative of the reaction to most of the individual announcements. The second cognitive bias is the 'Law of Small Numbers.' Tversky and Kahneman (1971) show that academics place too much confidence in the results of small samples, which Hogarth (p. 219) notes "are likely to be atypical." Accounting researchers are likely to fall prey to the 'Law of Small Numbers' by generalizing results from a small number of published research studies (often only a single study), which may have found atypical results.

The effects of these cognitive biases are exacerbated by the second set of factors — biases inherent in the review process. The review process in accounting suffers from a reluctance to publish replications and a prejudice against the null. Unlike the physical sciences, replications are rarely published in major accounting journals (except for the short-lived 'Improvements and Updates' section in *Contemporary Accounting Research*). Because of the limited publication of replications, accounting researchers have little

choice but to generalize from a small number of published studies. The documented prejudice against the null raises a concern whether studies that are published have ‘typical’ results. Greenwald (1975) points out that the majority of published research rejects the null hypothesis, and Lindsay (1994) reaches a similar conclusion after examining a subset of management accounting research. Greenwald argues that prejudice against the null delays the acquisition of knowledge by fostering the publication of studies whose results are true, but of limited generalizability.³⁸ In combination with the bias against replication, this can lead to a situation where the first published studies are more likely to reject the null, and these initial studies have a disproportionate effect on subsequent research due to the bias against publishing replications. In the context of information content research, such a bias suggests that any studies that failed to find information content (i.e. failed to reject the null hypothesis) would have faced more difficulty in the review process than the original papers that rejected the null hypothesis.

The third factor is a bias suffered by collective academic communities. Normal science is characterized by a tendency to overlook inconvenient evidence that is inconsistent with prevailing beliefs. Research results that do not turn out as expected are viewed as “just a research failure, one which reflects not on nature but on the scientist” (Kuhn, 1970, p. 35). Researchers interested in career advancement may hesitate to risk exile from the paradigm by questioning its core beliefs. This discounting of anomalous evidence combines with the ‘Law of Small Numbers’ and the reluctance to publish replications to foster an environment where researchers heavily weight the results of a small number of initial studies. If those initial results do not ‘tell the whole story’ (perhaps because of the prejudice against the null), these biases can delay the development, publication, and

dissemination of evidence that tells ‘the rest of the story’.

Beaver’s focus on cross-sectional mean reactions and the directed sampling approach (focusing on smaller firms with less predisclosure information) were reasonable research design choices — especially for the pioneering study in the historical context when the research was conducted. Nevertheless, if the initial information content studies had assessed the proportion of earnings announcements that were ‘informative’ (rather than focussing on the cross-sectional mean reaction), or if the samples had not consisted of smaller-than-average firms [e.g., non-12/31 NYSE firms with less than 20 news announcements in Beaver (1968) and AMEX firms in May (1971)], early views concerning the information content of earnings announcements would have been less sanguine. If subsequent researchers had more carefully examined the initial studies’ methods and evidence, if they had understood the implications of the initial studies’ sample selection criteria, or if they had appreciated that the mean market reaction might not be representative of ‘typical’ market reactions, the community of accounting researchers would have been more skeptical from the outset, and less likely to dismiss anomalous evidence (e.g. Oppong 1980). Our understanding of the earnings-return relation would likely have progressed more rapidly.

For example, financial accounting researchers would likely have focussed on *why* most earnings announcements do not convey ‘new news’ to the market. There are at least two possible explanations. Many firm-years’ earnings announcements: (1) are “too late” in that their information has been preempted, or (2) are not value-relevant. Investigation of the first explanation would have accelerated research on predisclosure information. Researchers suspecting that earnings announcements are too late to convey new information to the market would have investigated the extent to which other disclosures (e.g. interim earnings, analysts’ forecasts, voluntary management forecasts of earnings or sales) are informative and preempt the information content of the earnings announcement. Researchers may also have been more apt to recognize that related firms’ earnings

³⁸ At the extreme, with a 5% level of significance, 5% of studies will inappropriately reject the null. The prejudice against the null suggests that there may be a bias in favor of publishing the study that inappropriately rejects the null, rather than the majority that failed to reject the null.

announcements can pre-empt information in a late-announcing firm's earnings number (e.g. intra-industry information transfers).

Investigation of the second explanation would have hastened research on the value-relevance of earnings. The weakness of the earnings-return relation would have been understood prior to the late 1980's (e.g. Lev, 1989). Researchers would also have investigated cross-sectional and cross-temporal variation in the value-relevance of earnings, and begun much earlier the search for financial and non-financial firm value-drivers (e.g. Amir & Lev 1996; Barth, Clement, Foster & Kasznik 1998; see also recent management accounting research reviewed in Ittner & Larcker 1998a). Earlier documentation of the incremental value-relevance of non-financial information similar to Barth and McNichols (1994), Amir and Lev (1996), and Ittner and Larcker (1998b), coupled with greater insight into the apparent lack of new information content in most earnings announcements, might have stimulated a drive to expand the role of accounting beyond financial information well before 1994 when the Jenkins report was issued (AICPA, 1994). In sum, our collective failure to appreciate the effects of subjective research design choices, coupled with our misinterpretation of research evidence, delayed the acquisition of knowledge in ways that may have fundamentally affected the scope and role of accounting.³⁹

While it will not be simple for the community of accounting scholars to address the biases highlighted here, the first step is to increase awareness

that a problem exists, and this has been the primary objective of our study. More specifically, we recommend that:

1. Authors be forthright about the likely impact of research design choices on their results, and readers of research should carefully consider the likely effects of research design choices on the studies' results.⁴⁰
2. Readers and reviewers of research guard against falling prey to the 'Law of Small Numbers', and not generalize results based solely on one or two studies (because those studies may have found atypical results).
3. Authors, reviewers, and editors make a concerted effort to overcome prejudice against the null. Authors should carefully evaluate their research design choices to ensure that the analysis is sufficiently powerful to detect a relation, if one exists. Authors can increase the informativeness of statistically 'insignificant' results if they are able to construct confidence intervals that reveal the maximum effect is so small that it is economically insignificant. See Greenwald (1975) for further discussion on gracefully failing to reject the null, and Cready and Mynatt (1991) for a specific example in accounting.
4. Reviewers and editors should recognize the value of replications, and consider devoting some journal space to substantive replications

³⁹ Our consideration of how financial accounting research might have progressed had Beaver made different research design choices or had subsequent researchers not over-generalized his results, is necessarily speculative. However, this discussion is consistent with the relatively recent rise of counterfactual reasoning in social sciences including history (e.g. Ferguson, 1998) and political science (e.g. Fearon, 1991). According to Davis (1970, p. 76) "counterfactual arguments are a part of all but purely descriptive history." Philip Tetlock, a professor of psychology and political science, argues that "counterfactual reasoning is a prerequisite for any form of learning from history" (Orlans, 1997), because "history doesn't provide us with control groups, so we have to construct them in our imaginations. In order for us to learn anything from history, we have to think about what didn't happen along with what did." (*Psychology Today*, 1998).

⁴⁰ As an extreme example, Walster and Cleary (1970) propose a review process akin to that often applied to doctoral dissertations. They suggest that peer review occur before execution of the research, based on research proposals. Peer review exchanges between authors and referees could highlight likely effects of various research design choices before the work is executed. Walster and Cleary suggest that upon completion, research stemming from proposals approved by this peer review process could be published regardless of the statistical significance of the results. While their suggestion would certainly require close scrutiny prior to a consideration of implementation (for example, how could authors be motivated to carefully conduct the research, if publication were already guaranteed?), such a proposal might: (a) focus more attention on research design, (b) mitigate the bias against publishing insignificant results (consistent with our recommendation 3), and (c) prevent referees and editors from suppressing results that they do not like (consistent with our recommendation 5).

of important research. Unlike the hard sciences, replications have little stature in social sciences like accounting. This is particularly curious since our research consists largely of quasi-experiments using archival data that are often questionable proxies for the underlying constructs of interest, and laboratory experiments often based on participants who are questionable proxies for the population of decision-makers they are purported to represent.

5. When results of carefully conducted and powerful tests turn out unexpectedly, the entire community of accounting researchers should be on guard against blaming the scientist rather than nature.

Acknowledgements

We appreciate the helpful comments of Paul Ashcroft, Michael Bamber, Russ Barefield, Orie Barron, George Benston, Youngsoon Susan Cheon, Kirsten Ely, Tim Fogarty, George Foster, Robert Freeman, Julia Grant, Byron Henry, Anthony Hopwood, Keith Houghton, Skip Hughes, David Hurtt, Kimberly Galligan Key, Ray King, Robert Magee, Dana Northcut, Elizabeth Plummer, Gary Previts, Vaughn Radcliffe, Jerry Salamon, Katherine Schipper, Jim Seida, Brian Shapiro, Mary Stokes, Pamela Stuerke, Senyo Tse, Siddharta Utama, two anonymous reviewers, and workshop participants at the University of Arizona, Case Western Reserve University, University of Connecticut, Emory University, University of Georgia, University of Melbourne, Michigan State University, University of Oklahoma, University of Oregon, and Texas A&M University.

Appendix: Construction of the 'Beaver-like' sample

Exact reconstruction of Beaver's sample is impossible because the CRSP and COMPUSTAT tapes he used are not currently available. His CRSP tape included data from at least 1961,

whereas current CRSP tapes begin in July 1962. COMPUSTAT has a well-known survivorship bias in that it includes only firms existing when the tape is cut (i.e. it purges firms covered in previous years that subsequently went out of existence). Also, COMPUSTAT has extensively backfilled data (i.e. when COMPUSTAT begins covering a firm, it fills in historical as well as current period data). These practices considerably complicate reconstruction of Beaver's sample.

Our reconstruction is based on three assumptions. First, since Beaver's sample period covered 1961–1965, we assume that Beaver's COMPUSTAT tape ended with fiscal year 1965 data. Thus, we assume that his COMPUSTAT tape was cut between mid-1966 and mid-1967. Second, we assume that all firms in Beaver's sample were still actively traded on the NYSE in December 1965. Third, we assume that firms with data available on Beaver's CRSP tapes are those for which 1962–1966 data are available on the 1993 CRSP tapes. (This assumption reflects the fact that CRSP does not have a survivorship problem, and it also assumes that CRSP has not engaged in backfilling data. Discussions with CRSP representatives suggest that CRSP has not engaged in backfilling beyond minor corrections/completions.)

Data sources available to us included: (1) a COMPUSTAT quarterly back data tape covering 1960–1971, (2) an Annual Industrial COMPUSTAT tape covering 1961–1980, (3) a printed list of 1705 firms included on COMPUSTAT's Annual Industrial tape as of 1971, (4) ISL Guides to Stocks listed on the NYSE, AMEX, and OTC exchanges, and (5) the *Wall Street Journal Index*. Our sample reconstruction procedure consisted of six steps:

1. Select all firms on the 1993 quarterly back data tape that had a non-missing fiscal year-end value for fiscal year 1965. 846 of the 2,812 firms listed on the tape passed this screen. However, a limitation of this data source is that firms included on COMPUSTAT's Annual Industrial file during 1961–1965 may not have been covered on the quarterly files during the same time period. Thus, this screen does not recover all firms

that would have been available to Beaver. To address this problem, we performed steps 2 and 5b.⁴¹

2. Select all firms from the 1981 COMPUSTAT Annual Industrial tape that had a non-missing fiscal year-end date for 1965. 1,691 of the 2,457 firms listed on the tape met this criterion. A limitation of this data source is that this tape was cut in 1981, 14 to 15 years after Beaver conducted his study. Numerous firms were added to and deleted from the COMPUSTAT Annual Industrial file throughout this period. We perform step 5 below to address these backfilling and survivorship problems.
3. Merge the two sets of firms identified in steps 1 and 2 to obtain the sampling frame of firms from which our Beaver-like sample is selected. Many firms appear in both sets of data, and so merging the two datasets identifies 1,976 candidate firms for our Beaver-like sample.⁴²
4. From the 1,976 candidate firms identified in step 3, delete all firms that had 12/31 fiscal year-end dates in 1965. This screen reduced our sample by 68% to 624 firms. This percentage reduction is consistent with that Beaver reports in his Table 1, where he finds that application of the 12/31 screen reduced his sample by 67%.
5. Adjust for backfilling and survivorship problems.

5(A) Of the 624 firms from step 4, identify those that also appear on the list of

firms covered by the 1971 COMPUSTAT Annual Industrial tape. This step eliminates firms that were backfilled between 1971 and 1993. 464 firms pass this screen.⁴³

- 5(B) Identify firms that (1) appear on our printed list of firms covered by the 1971 Annual Industrial tape, and that (2) are not among the 464 firms from step 5(A), and that (3) had non-12/31 fiscal year end dates in 1965. This step corrects for survivorship bias. It recovers firms for which COMPUSTAT had in fact reported annual data in 1971, but which were eliminated from subsequent COMPUSTAT files. This step yielded an additional 56 firms.⁴⁴
- 5(C) Pool the two subsets of firms from steps 5(A) and 5(B) to yield 520 firms that (1) have a non-12/31 1965 fiscal year-end date per COMPUSTAT, (2) are known to have been on COMPUSTAT as of 1971, and (3) are highly likely to have been on COMPUSTAT during 1965.

6. From the 520 firms identified in step 5(C), eliminate all firms that were not traded on the NYSE in 1965. Also, eliminate any firms whose fiscal year end date was actually 12/31, per the *Wall Street Journal Index*. This yielded a sample of 255 firms for our Beaver-like sample, which is very close to the 242 non-12/31 NYSE firms Beaver identified in his Table 1.⁴⁵ Despite our access to very different data resources, our sample selection yielded a sample size within 5% of Beaver's. Thus, our sample is a good approximation of that used by Beaver.

⁴¹ It is unlikely that any firms on the quarterly file were not covered on the annual file.

⁴² Beaver notes in his Table 1 that he found 896 firms on his COMPUSTAT tape. Our step 3 reveals that in later years, COMPUSTAT has provided fiscal year 1965 data for 1,976 firms. This evidences COMPUSTAT's extensive backfilling between 1967 and 1993. Our step 6 suggests that much of this backfilling involved non-NYSE firms.

⁴³ These 464 firms may include some firms that were backfilled between 1967 and 1971. Such firms would not have been available to Beaver. Unfortunately, we have no means of identifying such firms. However, the results of step 6 below suggest that this is unlikely to be a serious problem.

⁴⁴ These recovered firms must be firms for which COMPUSTAT reported annual but not quarterly data, or else the firms would have appeared on the quarterly back data tape in step 1.

⁴⁵ Our NYSE screen eliminated about 50% of the sample firms from step 5c, which is considerably more than the approximately 20% eliminated by this screen in Beaver's sample (per his Table 1). This difference suggests that much of COMPUSTAT's backfilling efforts were directed toward non-NYSE firms.

References

- American Accounting Association (1977). *Statement on accounting theory and theory acceptance*. Sarasota, FL: American Accounting Association.
- American Institute of Certified Public Accountants (1994). *Improving Business Reporting — A Customer Focus: Meeting the Needs of Investors and Creditors*. (Comprehensive Report of the Special Committee on Financial Reporting) New York, NY: AICPA.
- Amir, E., & Lev, B. (1996). Value-relevance of nonfinancial information: The wireless communications industry. *Journal of Accounting and Economics*, 22(August/December), 3–30.
- Ashley, J. (1962). Stock prices and changes in earnings and dividends: Some empirical results. *Journal of Political Economy*, 70(February), 82–85.
- Atiase, R. (1985). Predisclosure information, firm capitalization, and security price behavior around earnings announcements. *Journal of Accounting Research*, 23(Spring), 21–36.
- Ball, R. (1994). The development, accomplishments and limitations of the theory of stock market efficiency. *Managerial Finance*, 20, 3–48.
- Ball, R., & Brown, P. (1968). An empirical evaluation of accounting income numbers. *Journal of Accounting Research*, 6(Autumn), 159–178.
- Bamber, L. (1987). Unexpected earnings, firm size, and trading volume around quarterly earnings announcements. *The Accounting Review*, 62(July), 510–532.
- Barth, M., Clement, M., Foster, G., & Kasznik, R. (1998). Brand values and capital market valuation. *Review of Accounting Studies*, 3, 41–68.
- Barth, M., & McNichols, M. (1994). Estimation and market valuation of environmental liabilities relating to Superfund sites. *Journal of Accounting Research*, 32(Suppl.), 177–209.
- Beaver, W. (1968). The information content of annual earnings announcements. *Empirical Research in Accounting: Selected Studies*, 1968, Supplement to *Journal of Accounting Research*, 6, 67–92.
- Beaver, W. (1989). *Financial reporting: An accounting revolution* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Bedford, N. (1957). Accounting measurements of economic concepts. *The Journal of Accountancy*, 103(May), 56–62.
- Benston, G. (1967). Published corporate accounting data and stock prices. *Empirical Research in Accounting: Selected Studies*, 1967, Supplement to *Journal of Accounting Research*, 1–14, 22–54.
- Bernard, V. (1989). Capital markets research in accounting during the 1980's: A critical review. In T. Frecka, *The State of accounting research as we enter the 1990's* (pp. 72–120). Urbana-Champaign, IL: University of Illinois.
- Bernard, V. (1991). Stock price reactions to earnings announcements: A summary of recent anomalous evidence and possible explanations. In R. Thaler, *Advances in behavioral finance* (pp. 303–340). New York: Russell Sage Foundation.
- Blough, C. (1937). How companies may make statements more illuminating to stockholders. *The Controller*. Reprinted in: W. Shenkir, Carman G. Blough: *His professional career and accounting thought* (pp. 282–285). New York: Arno Press.
- Brown, L. (1996). Influential accounting articles, individuals, Ph.D. granting institutions, and faculties: A citational analysis. *Accounting, Organizations and Society*, 21(October/November), 723–754.
- Brown, P. (1989). Ball and Brown [1968]. *Journal of Accounting Research*, 27(Suppl.), 202–217.
- Burgstahler, D. (1987). Inference from empirical research. *The Accounting Review*, 62(January), 203–214.
- Canning, J. (1929). *The economics of accountancy: A critical analysis of accounting theory*. New York: The Ronald Press Company.
- Chambers, R. (1974). *Accounting, evaluation and economic behavior*. Houston, TX: Scholars Book Co.
- Cready, W., & Mynatt, P. (1991). The information content of annual reports: A price and trading response analysis. *The Accounting Review*, 66(April), 291–312.
- Davidson, J. (1968). Discussion of: The information content of annual earnings announcements. *Empirical Research in Accounting: Selected Studies*, 1968, Supplement to *Journal of Accounting Research*, 6, 96–100.
- Davis, L. (1970). And it will never be literature: The new economic history. A critique. In R. Andreano, *The new economic history: Recent papers on methodology* (pp. 67–83). New York: John Wiley and Sons.
- Dyckman, T., Downes, D., & Magee, R. (1975). *Efficient capital markets and accounting: A critical analysis*. Englewood Cliffs, NJ: Prentice-Hall.
- Dyckman, T., & Morse, D. (1986). *Efficient capital markets and accounting: A critical analysis*. (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Fama, E., Fisher, L., Jensen, M., & Roll, R. (1969). The adjustment of stock prices to new information. *International Economic Review*, 10(February), 1–21.
- Fearon, J. (1991). Counterfactuals and hypothesis testing in political science. *World Politics*, 43(January), 169–195.
- Ferguson, N. (1998). *Virtual history: Alternatives and counterfactuals*. London: Papermac.
- Forbes. (1967). What are earnings? The growing credibility gap. *Forbes* (May 15), 28–31.
- Foster, G. (1986). *Financial statement analysis*. Englewood Cliffs, NJ: Prentice-Hall.
- Gilman, S. (1939). *Accounting concepts of profit*. New York: The Ronald Press Company.
- Grant, E. (1980). Market implications of differential amounts of interim information. *Journal of Accounting Research*, 18(Spring), 255–268.
- Greenwald, A. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin*, 82(January), 1–20.
- Greer, H. (1938). Application of accounting rules and standards to financial statements. *The Accounting Review*, 13(December), 333–345.
- Griffin, P. (1987). *Usefulness to investors and creditors of information provided by financial reporting*. Stamford CT: Financial Accounting Standards Board.
- Hogarth, R. (1987). *Judgement and choice: The psychology of decision* (2nd ed.). New York, NY: John Wiley and Sons.

- Ittner, C., & Larcker, D. (1998a). Innovations in performance measurement: Trends and research implications. *Journal of Management Accounting Research*, 10, 205–238.
- Ittner, C., & Larcker, D. (1998b). Are nonfinancial measures leading indicators of financial performance? An analysis of customer satisfaction. *Journal of Accounting Research*, 36(Suppl.), 1–35.
- Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgement of representativeness. *Cognitive Psychology*, 3(July), 430–454.
- Kaplan, R. (1978). The information content of financial accounting numbers: A survey of empirical evidence. In R. Abdel-khalik, & T. Keller, *The impact of accounting research on practice and disclosure* (pp. 134–173). Durham, NC: Duke University Press.
- Kuhn, T. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago, IL: University of Chicago Press.
- Leamer, E. (1983). Let's take the con out of econometrics. *The American Economic Review*, 73(March), 31–43.
- Lev, B. (1989). On the usefulness of earnings and earnings research: Lessons and directions from two decades of empirical research. *Journal of Accounting Research*, 27(Suppl.), 153–192.
- Lindsay, R. (1994). Publication system biases associated with the statistical testing paradigm. *Contemporary Accounting Research*, 11(Summer), 33–57.
- Marais, M. (1984). An application of the bootstrap method to the analysis of squared, standardized market model prediction errors. *Journal of Accounting Research*, 22(Suppl.), 34–54.
- May, R. (1971). The influence of quarterly earnings announcements on investor decisions as reflected in common stock price changes. *Empirical Research in Accounting: Selected Studies*, 1971. Supplement to *Journal of Accounting Research*, 119–163.
- Mouck, T. (1993). The “revolution” in financial accounting theory: A Kuhnian interpretation. *The Accounting Historians Journal*, 20(June), 33–57.
- Mouck, T. (1998). Capital markets research and real world complexity: The emerging challenge of chaos theory. *Accounting, Organizations and Society*, 23(February), 189–215.
- Oppong, A. (1980). Information content of annual earnings announcements revisited. *Journal of Accounting Research*, 18(Autumn), 574–584.
- Orlans, H. (1997). If Bosnians were dolphins. *Change*, 29(May), 7–8.
- Parker, C. (1967). Discussion of: Published corporate accounting data and stock prices. *Empirical Research in Accounting: Selected Studies*, 1967, Supplement to *Journal of Accounting Research*, 5, 15–18.
- Patell, J. (1976). Corporate forecasts of earnings per share and stock price behavior: Empirical tests. *Journal of Accounting Research*, 14(Autumn), 246–276.
- Paton, W., & Littleton, A. (1940). *An introduction to corporate accounting standards*. (Reprinted in 1977). American Accounting Association, Monograph No. 3.
- Psychology Today. (1998). What if you hadn't read this... Vol. 31, (September), 12.
- Shores, D. (1990). The association between interim information and security returns surrounding earnings announcements. *Journal of Accounting Research*, 28(Spring), 164–181.
- Smith, D., & Pourciau, S. (1988). A comparison of the financial characteristics of December and non-December year-end companies. *Journal of Accounting and Economics*, 10(December), 335–344.
- Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, 76(August), 105–110.
- Vatter, W. (1947). *The fund theory of accounting and its implications for financial reports*. Chicago, IL: The University of Chicago Press.
- Wallace, W. (1991). *Accounting research methods: Do the facts speak for themselves?* Homewood, IL: Richard D. Irwin.
- Walster, G., & Cleary, T. (1970). A proposal for a new editorial policy in the social sciences. *The American Statistician* (April), 16–19.
- Watts, R., & Zimmerman, J. (1986). *Positive accounting theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Wells, M. C. (1976). A revolution in accounting thought? *The Accounting Review*, 51(July), 471–482.