APUBEF

Pennsylvania Journal of Business and Economics
Volume 12 Number 1 ISSN: 2156-9444 Fall 2006

PJBE Office of the Editor

Disclaimer
The facts, opinions, and conclusions set forth in the articles contained in this issue are those of the authors, and quotations should be so attributed. They do not necessarily represent the views and opinions of the Association of Pennsylvania University Business and Economics Faculty (APUBEF) or the editors; nor can APUBEF and the editors assume any responsibility for the accuracy or validity of any of the information contained herein.

EDITORIAL NOTES

The Fall 2006 edition of the *Pennsylvania Journal of Business and Economics* (PJBE) is the culmination of the efforts of many individuals who volunteered their time and energy to create a quality, general-interest business and economics journal that is both useful and enjoyable to read. We thank the members of the Editorial Review Board and their colleagues who agreed to review manuscripts for this edition.

All manuscripts accepted for publication in this edition underwent both a double-blind review for content and a rigorous review of grammar, formatting and style. We thank all the authors for their patience with this process; it can get lengthy at times. However, we believe it is worth the extra time and effort to help produce a quality, well-written journal. The acceptance rate for this issue is once again 40%. This is a continuation of our reviewers' efforts to increase standards, while still providing useful feedback to authors.

The PJBE continues to be listed in Cabell's, and all information about the journal was recently updated. Dr. Leon Markowicz continues to handle the review process for this edition. Dr. Kevin Roth continues with editing, production, and distribution of the journal.

Finally, we thank all those individuals who submitted manuscripts for possible inclusion in this edition. We encourage all our colleagues to consider the PJBE as an outlet for their work.

Table of Contents

IMPROVING PROFESSIONAL AUDITORS' GOING-CONCERN JUDGMENTS

Kim L. Anderson, Indiana University of Pennsylvania

ABSTRACT

This paper examines the effects of "hindsight bias" on professional auditors' going-concern judgments and the degree to which the bias is mitigated by a preoutcome debiasing strategy found to be successful in the psychological literature. Hindsight bias is the tendency for individuals who have been provided the outcome of an event to overstate their abilities to have predicted that outcome in foresight. Using an experimental methodology, the results indicate that a preoutcome debiasing strategy (referring back to reasons for the alternative outcomes recorded in foresight) produces asymmetrical effects in an audit setting involving going-concern judgments made by professional auditors employed by international public accounting firms. Auditors record a greater number of more highly rated reasons supporting the failure outcome as compared to the success outcome. Referring back to the reasons eliminates hindsight bias for the success outcome, but increases the bias for the failure outcome.

INTRODUCTION

This paper examines the effects of "hindsight bias" on auditors' going-concern judgments and the degree to which the bias is mitigated by a debiasing strategy found to be successful in the psychological literature. The subjects used in the current study are professional auditors employed by an international public accounting firm. Hindsight bias is the tendency for individuals who have been provided with the outcome of an uncertain event to systematically overstate their abilities to have predicted that outcome in foresight. Further, individuals deny that knowledge of the event's actual outcome has affected their predictions. Hindsight bias has been found to influence several audit judgments, including internal control evaluations (Reimers & Butler, 1992), audit opinion decisions (Reimers & Butler, 1992), preliminary analytical review judgments (Biggs & Wild, 1985; Heintz & White, 1989; Kennedy, 1995; Kinney & Uecker, 1982; McDaniel & Kinney, 1994), and going-concern judgments (Anderson, 2000, 2002; Kennedy, 1993, 1995; Maddocks, 1989).

According to Fischhoff (1975), the "knew-it-all-along" attitude created by hindsight bias impedes feedback learning, thereby reducing what individuals could potentially learn from the feedback provided by the outcome. If auditors believe they "knew all along" that a bankrupt company was going to fail, they will not learn what they should from the outcome and will believe more often than they should that they could have actually predicted the outcome. This overconfidence may lead auditors to believe they have little reason to re-evaluate and improve their decision making processes and evidence gathering strategies regarding going-concern judgments. [1]

In light of the recent spate of U.S. corporate bankruptcies and audit failures, it is more important than ever for auditors to improve their going-concern judgments by learning from the feedback provided by actual bankruptcies. Of the 15 largest U.S. corporate bankruptcies in the past 25 years, all but five occurred after the first quarter of 2001, including the two largest corporate bankruptcies in history, Worldcom, Inc. and Enron Corporation (BankrupcyData.com, 2006). Case studies describing the facts surrounding many of these bankruptcies are appearing in auditing textbooks and are being used in public accounting firm training programs (Ahrens et al., 2006). In order to learn as much as possible from the feedback provided by these case studies, it is imperative that auditing students, as well as professional auditors, be provided with debiasing strategies designed to reduce hindsight bias.

In an auditing experiment involving going-concern judgments, this paper investigates the effectiveness of a debiasing strategy found to be successful in reducing hindsight bias in the psychological literature. Given that monetary incentives (Camerer et al., 1989; Hell et al., 1988), accountability (Kennedy, 1993, 1995), and experience (Anderson, 2000; Kennedy, 1995) have been found ineffective in counteracting hindsight bias, it is important to discover decision aids that are successful in reducing it.

In a series of three psychological experiments, Davies (1987) found that hindsight bias is reduced by allowing subjects to review reasons for the alternative outcomes that they had recorded before they were informed of the actual outcome. Davies theorized that this preoutcome debiasing strategy enables subjects to retrieve their foresight perspectives. Davies also found that hindsight bias is reduced by instructing subjects to record reasons for the alternative outcomes after they were informed of the actual outcome. Davies argued that this postoutcome debiasing strategy redresses the imbalance between the greater availability of reasons favoring the reported outcome as compared to the nonreported outcome. Davies found both of these debiasing strategies to be equally effective. However, according to Davies, for events whose occurrence and importance are known in advance (such as elections, space launches, and impending bankruptcies), the preoutcome debiasing strategy may be more useful.

In an auditing study involving an analytical review task, Kennedy (1995) found that a debiasing strategy similar to Davies' postoutcome debiasing strategy is successful in an auditing context. In an experiment involving auditors' going-concern judgments, this paper examines the extent to which Davies' preoutcome debiasing strategy is successful in an auditing context. Due to the unique nature of auditors' training and experience, the current study predicts and finds that, when instructed to generate reasons for the alternative outcomes in foresight, an asymmetrical effect is produced, whereby auditors self-generate a greater number of more highly rated reasons supporting the failure outcome as compared to the success outcome. As a result, referring back to the lists after the receipt of outcome information eliminates the degree of hindsight bias exhibited by auditors provided with the success outcome. However, reviewing the lists substantially increases the bias for auditors provided with the failure outcome by creating a "I-really-did-know-it-all-along" attitude.

The main contribution of this study is that debiasing strategies found to be successful in reducing hindsight bias in the psychological literature may produce asymmetrical effects in an auditing environment by eliminating the bias for some outcomes and by exacerbating it for others. Due to auditors' unique training and experience, it cannot be assumed that auditors will behave and respond in the same manner as subjects in psychological experiments. This illustrates the need to exercise caution when importing results from psychological literature to an auditing domain. It may first be necessary to subject the findings in the psychological literature to empirical testing that includes auditor subjects performing auditing tasks.

THEORY AND HYPOTHESES DEVELOPMENT

Presence of hindsight bias

Fischhoff (1975) coined the term "creeping determinism" to describe the process he believed was responsible for hindsight bias. According to Fischhoff, "Upon receipt of outcome knowledge judges immediately assimilate it with what they already know about the event in question. In other words, the retrospective judge attempts to make sense, or a coherent whole, out of all that he knows about the event" (1975, 297). Because the process was hypothesized to be quick and unconscious, Fischhoff described the outcome information as "creeping" into the subject's mental representation of the event resulting in cognitive restructuring. The characteristic effect of creeping determinism is the proclivity to view a known outcome as nearly inevitable, as revealed in retrospective probability judgments, because of the seemingly unalterable sequence of events leading up to it (Hawkins & Hastie, 1990). The "creeping determinism" hypothesis is consistent with more of the hindsight literature results than any other explanation offered (Hawkins & Hastie, 1990). [2]

Prior research reveals the presence of hindsight bias in several accounting settings. Financial statement users asked to assess a company's viability have been found to be prone to hindsight bias (Buchman,1985). Jurors (Lowe & Reckers, 1994) and judges (Anderson et al., 1995; Anderson et al., 1997) asked to evaluate the actions of auditors have also been found to be prone to the bias. Brown and Solomon (1987) found that capital-budgeting decisions are influenced by outcome information. In an auditing study involving internal control evaluations and audit opinion decisions, Reimers and Butler (1992) found that auditors exhibit significant (insignificant) hindsight bias when provided with surprising (unsurprising) outcome information. Anderson (2000, 2002) and Kennedy (1993, 1995) found that auditors are prone to hindsight bias when making going-concern judgments, and Kennedy (1995) found that auditors exhibit the bias when making analytical review judgments.

These findings suggest that auditors are prone to hindsight bias. As a result, the current study predicts, despite instructions to ignore outcome information, auditors provided with outcome information will exhibit hindsight bias when making going-concern judgments. This leads to the following baseline hypothesis:

> H1: Despite instructions to ignore outcome information, auditors with outcome information will judge the reported outcome as more likely to occur than will auditors not provided with outcome information.

Failure outcome versus success outcome

According to the psychological literature, an occurrence results in greater hindsight bias than does a nonoccurrence (Fischhoff, 1977; Fischhoff & Beyth, 1975; Wasserman et al., 1991; Wood, 1978). A nonoccurrence results in lower hindsight bias because it is regarded as a nonevent which requires very little cognitive restructuring (Fischhoff, 1977; Schkade & Kilbourne, 1991). An occurrence, on the other hand, results in substantial cognitive restructuring and, therefore, greater hindsight bias.

As Kennedy (1995) points out, in a going-concern task, it is likely subjects would regard success as the nonoccurrence because it is a continuation of the status quo. Failure, on the other hand, would be viewed as an interruption of the status quo, as an occurrence. In an experiment involving a going-concern task, Kennedy (1995) did find that auditors exhibit greater hindsight bias when informed of a failure outcome as compared to a success outcome. This leads to the following baseline hypothesis:

> H2: Auditors informed of a failure outcome will exhibit greater hindsight bias than will auditors informed of a success outcome.

Reducing hindsight bias

Attempts to eliminate or even reduce hindsight bias have been only moderately effective. Exhorting subjects to work hard and cautioning them about the bias have been ineffective (Fischoff, 1982; Wood, 1978). Hasher et al. (1981) were successful in eliminating hindsight bias, but only by discrediting the outcome information in such a manner that subjects realized that it was totally unreliable. Wood (1978) found that preoutcome judgments can be used to decrease hindsight bias, but only if subjects were encouraged to remember their previous judgments when making their postoutcome judgments. In an experimental markets study, Camerer et al. (1989) found that feedback and monetary incentives alone had no effect on reducing hindsight bias; however, market forces reduced the bias by approximately one half.

Instructing subjects to generate reasons for the alternative outcomes <u>after</u> the receipt of outcome information has been found to significantly reduce, but not eliminate, hindsight bias. This postoutcome debiasing strategy has been found to reduce hindsight bias in psychological studies involving student subjects (Davies, 1987; Slovic & Fischhoff, 1977), in a medical diagnosis study involving physicians (Arkes et al., 1988), and in an accounting study involving jurors' evaluations of auditors' decisions (Lowe & Reckers, 1994). It has been theorized (Davies, 1987; Slovic & Fischhoff, 1977) that a postoutcome debiasing strategy reduces hindsight bias by focusing the subjects' attention on the nonreported outcome and away from the reported outcome. This adjusts the imbalance between the greater availability of reasons favoring the reported outcome as compared to the nonreported outcome. In an experiment involving an analytical review task, Kennedy (1995) found that a postoutcome debiasing strategy (instructing auditors to generate reasons for the nonreported outcome) eliminated hindsight bias.

Instructing subjects to generate reasons for the alternative outcomes <u>before</u> the receipt of outcome information has also been found to significantly reduce, but not eliminate, hindsight bias. This preoutcome debiasing strategy has been found to be successful in reducing hindsight bias both when the subjects were allowed to refer back to their recorded reasons (Davies, 1987) and when they were not allowed to refer back (Hell et al., 1988). It has been theorized that a preoutcome debiasing strategy enables the subjects to recover their foresight perspectives, which, in turn, counteracts the effects of creeping determinism, thereby reducing hindsight bias (Davies, 1987). The current study examines the effectiveness of a preoutcome debiasing strategy similar to Davies' (1987).

Although Kennedy (1995) has already found that a postoutcome debiasing strategy is effective in eliminating hindsight bias in an audit setting, it is important to also examine the effectiveness of a preoutcome debiasing strategy for two main reasons. First, although Davies (1987) found preoutcome and postoutcome debiasing strategies to be equally effective, he also points out that, for events whose occurrence and importance are known in advance (such as elections, space launches, and impending bankruptcies), a preoutcome debiasing strategy may be more useful. Davies argues that in situations involving experts (e.g., auditors), "High involvement and expertise might lead the foresightful judge to produce more detailed and more organized records of their foresight state of knowledge that would then serve as more powerful retrieval cues in hindsight" (1987, 66). Second, given that a preoutcome debiasing strategy may be more useful in certain situations, it is important to discover to what extent it is indeed more effective in an audit setting. Although Davies argues that experts may be able to record more detailed and organized foresight records, he did not include experts when testing his preoutcome debiasing strategy. As discussed in the next section, Davies' preoutcome debiasing strategy may be only partially effective in an audit setting involving going-concern judgments.

Davies' subjects were undergraduate students, and the nature of their experimental task was to review case data describing a simple psychological experiment that had two possible outcomes. The subjects were not psychologists and were not familiar with the information presented in the case. The subjects had no prior training or experience that might have caused them to view one of the possible outcome as more important or to view one of the possible outcomes as more rare or unusual than the other. As a result, there was no reason to expect the subjects, as a group, to record significantly more information about one of the outcomes or to rate the reasons listed for one outcome as more relevant than the reasons listed for the alternative outcome.

The subjects in the current study, however, were professional auditors, all of whom had some level of previous training in making going-concern judgments. Auditors are trained to classify a company as one with going-concern problems only if, during normal auditing procedures, sufficient adverse factors have been identified to create substantial doubt about the company's ability to continue as a going concern. Only after sufficient adverse factors have been uncovered does the auditor search for mitigating factors (AICPA, 1988). This may train auditors to be more aware of adverse factors and to view adverse factors as more important.

Other issues may contribute to the tendency for auditors to be more aware of adverse factors and to view adverse factors as more important. First, the failure to uncover an existing adverse factor poses more dire consequences to the public accounting firm than does the failure to uncover a mitigating factor. Second, client management may be inclined to conceal adverse factors from the auditors, but has little incentive to conceal mitigating factors. Third, the failure outcome is a rare event which may cause auditors to view the reasons (i.e., adverse factors) leading up to it as more salient than the reasons (i.e., mitigating factors) leading up to a normally occurring event such as the success outcome. As a result, when asked to generate a list of adverse factors, auditors are likely to list significantly more adverse factors and to rate the adverse factors as significantly more important.

This is consistent with Kida's (1984) findings in a study designed to test auditor's hypothesis-testing strategies. Kida found that audit partners and managers, who were asked to list the most relevant adverse factors and mitigating factors from the case description of a troubled company, recorded significantly more adverse factors. This leads to the following hypothesis:

> H3: When asked to self-generate a list of adverse factors and mitigating factors prior to the receipt of outcome information, auditors will record more adverse factors than mitigating factors and will rate the adverse factors as more important.

Debiasing strategy and success outcome

When auditors who been provided with success outcome information refer back to their lists of adverse factors and mitigating factors that they recorded in foresight, they will review a list which contains a greater number of more highly rated adverse factors, factors which support the nonreported outcome. Consistent with Davies (1987), this may help to counteract the creeping determinism caused by the success outcome information, thereby enabling the auditors to recapture their foresight perspectives. This, in turn, should reduce the degree of hindsight bias exhibited. This is hypothesized more formally as follows:

> H4a: Allowing auditors who have outcome information to review their lists of adverse factors and mitigating factors that were recorded in foresight will reduce hindsight bias for auditors with success outcome information.

Debiasing strategy and failure outcome

The greater number of adverse factors recorded in conjunction with their higher importance ratings will lead to increased hindsight bias for auditors provided with failure outcome information. Upon receipt of the failure outcome information, the auditors will experience creeping determinism, whereby the failure outcome will alter their mental representations of the case scenario causing the failure outcome to appear virtually inevitable in hindsight. When asked to ignore the failure outcome and to refer back to their lists of adverse factors and mitigating factors in an effort to make the going-concern judgment as they would have in foresight, the lists of factors, which will contain a greater number of more highly rated adverse factors, will likely cause the opposite of the desired effect. Rather than aiding the auditors in recapturing their foresight states of uncertainty, the lists of factors may confirm and enhance their hindsight states of certainty by invoking a "I-really-did-know-it-all-along" reaction.

This increased hindsight bias did not occur in Davies' (1987) study because his subjects did not record, as a group, significantly more reasons for one of the two possible outcomes and had no previous training compelling them to view one of the outcomes as more important. Upon receipt of outcome information, when Davies' subjects referred back to their lists of foresight reasons, they were not faced with lists that strongly supported the reported outcome. Instead, their lists contained a more balanced number of reasons supporting both the reported and the nonreported outcomes, thereby enabling them to recapture their foresight states of uncertainty rather than bolstering their hindsight states of certainty. Based on the foregoing, the following hypothesis is proposed:

> H4b: Allowing auditors who have outcome information to review their lists of adverse factors and mitigating factors that were recorded in foresight will increase hindsight bias for auditors with failure outcome information.

In summary, consistent with prior auditing research (Kennedy, 1993, 1995; Reimers & Butler, 1992), the current study predicts that auditors will be prone to hindsight bias when making going-concern judgments. Also consistent with prior auditing research (Kennedy, 1995), the degree of hindsight bias exhibited is predicted to be greater for auditors provided with failure outcome information as compared to success outcome information. It is also predicted that, due to the unique nature of audit training and experience, a preoutcome debiasing strategy found to be successful in the psychological literature (Davies, 1987) will produce asymmetric effects in an audit setting by reducing hindsight bias in the case of the success outcome, but increasing it in the case of the failure outcome. The experiment designed to test the hypotheses is described next.

RESEARCH METHOD

Experimental design

In order to test the proposed hypotheses, one experiment was conducted. The basic design used is a 2X3 factorial. The two between factors are debiasing strategy and outcome. The debiasing strategy factor has two levels, no (i.e., the debiasing strategy was not used) and yes (i.e., the debiasing strategy was used). The outcome factor has three levels: no outcome, (i.e., the foresight condition), failure outcome (i.e., the hindsight condition – the occurrence of bankruptcy), and the success outcome (i.e., the hindsight condition – the nonoccurrence of bankruptcy). The dependent variable is the auditor's going-concern probability judgment (hereafter referred to as a viability judgment).

Subjects and procedure

The subjects were asked to judge the likelihood that a troubled company would or would not continue as a going concern. The sample of subjects consisted of 228 auditors from international public accounting firms. To obtain a sufficient number of subjects, it was necessary to administer the experiment at 14 different sessions over the course of four months. Responses to the debriefing questionnaire revealed that the mean auditing experience to be 5.4 years.

Subjects were randomly assigned to experimental conditions. Each subject received a packet of materials, consisting of a sealed envelope, a page of general instructions, and either five or six pages of case data (including a case review task). After completing the case review task, the written instructions indicated that the subjects were to open the sealed envelope. The envelope contained: the outcome information (if provided), the viability judgment task, and the debriefing task. The subjects were not allowed to use reference materials and were required to work independently.

Case review task

The subjects were provided with a page of general instructions. They also received a narrative summary of pertinent information for a chemical manufacturer and three years of financial data. The narrative summary contained an equal number of adverse factors and mitigating factors. The financial data included the financial statements, a summary of financial highlights, and a set of financial ratios.[3]

Figure 1 illustrates the experimental tasks that the subjects were asked to perform. The subjects' first task was to review the case data for Alpha Chemical, Inc. They were instructed to assume the role of supervisor on the Alpha audit for year 1. They were also told that the fieldwork had been completed, but the final audit opinion had not yet been written. They were to review Alpha's financial statements in an attempt to assess viability.

Figure 1
Experimental Tasks

	No Debiasing Strategy			Debiasing Strategy		
	No Outcomes	Failure Outcome	Success Outcome	No Outcome	Failure Outcome	Success Outcome
STEPS						
I.	Review Case Data (Task #1) No Writing			Review Case Data (Task #1) Writing -Prior to Receipt of Outcome Information -Instructed to Record Adverse & Mitigating Factors -Instructed to Rate Importance of Each Factor Revealed		
II.	No Outcome	Told Failure	Told Success	No Outcome	Told Failure	Told Success
III.	-Allowed to Refer to Case Data -Viability Judgment (Task #2)			-Allowed to Refer to Case Data -Instructed to Review Written Record of Factors and Importance Ratings -Viability Judgment (Task #2)		
IV.	Debriefing Task (#3)			Debriefing Task (Task #3)		

Debiasing strategy manipulation

While reviewing the case data and before receiving outcome information (if provided), subjects in the debiasing strategy group were instructed to record as many mitigating factors (i.e., reasons and information pointing toward continued success) and adverse factors (i.e., reasons and information pointing toward failure) as they could that they believed should be considered in determining Alpha's viability. They were encouraged to go beyond the case data and to rely on their existing knowledge and experience in recording the factors. They were also instructed to rate the importance of each adverse factor and mitigating factor they recorded. They were provided with a 4-point scale from 1, somewhat important, to 4, very important, and were asked to place either a 1, 2, 3, or 4 next to each factor they listed.

Viability judgment task

After reviewing the case data, subjects were instructed to begin the second task, the viability judgment. Before making their viability judgments, subjects in the failure outcome condition were informed that the company did file for bankruptcy during the last half of the year subsequent to the year under audit. Subjects in the success outcome condition were informed that the company did continue in existence as a going concern throughout the year subsequent to the year under audit. Subjects in the no outcome condition were not provided with any outcome information.

All subjects were instructed to assume that it was the last day of fieldwork for the year-end audit. They were reminded that, at that time, they would not have known whether the company had succeeded or failed, so were told to ignore the fact that they now know the outcome. They were instructed to estimate the likelihood that the company would or would not continue as a going concern throughout the year subsequent to the year under audit by placing an "X" on a probability scale ranging from 0% (certain NOT to continue) to 100% (certain to continue). [4]

Debriefing task

The final task for all subjects was completing a one-page debriefing questionnaire. Subjects were asked to indicate their number of years and months of experience, their current rank in their firm, and the number of minutes they took in completing the experiment. They were also asked to indicate both the number of audit engagements they had been associated with in which substantial doubt existed regarding the client's ability to continue as a going concern and their degree of involvement in the going-concern evaluation of these clients. In addition, they were asked to rate their degree of proficiency at evaluating a company's going-concern status. Finally, subjects in the failure outcome and success outcome conditions were asked to indicate the degree of influence, if any, the outcome information had on their viability judgments.

RESULTS

Results of hypothesis 1

H1 predicted that, despite instructions to ignore outcome information, auditors with outcome information would judge the reported outcome as more likely to occur than would auditors not provided with outcome information. More specifically, auditors informed that the case company failed (continued) would be more likely to judge the continued viability of the company as being less (more) likely than the auditors not provided with outcome information. The means and standard deviations for the viability judgment dependent variable are presented in Table 1. The viability judgment scale ranged from 0%, the company is certain not to continue, to 100%, the company is certain to continue.

Table 1
Means & (Standard Deviations) of Viability Judgments by Experimental Conditions

DEBIASING STRATEGY	Outcome: No	Outcome: Failure	Outcome: Success
No	60.66% (17.24) n = 38	54.37% (17.06) n = 38	67.00% (17.80) n = 38
Yes	64.53% (17.10) n = 38	45.13% (19.75) n = 38	68.53% (16.11) n = 38

ANOVA results – interaction effects

To test the effect of outcome information on auditors' viability judgments, a 2X3 (debiasing strategy by outcome) ANOVA was performed. The debiasing strategy factor has two levels (i.e., no debiasing strategy and debiasing strategy), and the outcome factor has three levels (i.e., no, failure, and success).

Table 2
ANOVA: Debiasing Strategy by Outcome on Viability Judgments

Source of Variation	SS	DF	MS	F	sig. of F
Strategy	.009	1	.009	.304	.582
Outcome	1.308	2	.654	21.229	.000
Strategy by Outcome	.186	2	.093	3.014	.051

The ANOVA results are presented in Table 2. The two-way interaction between debiasing strategy and outcome is significant ($p = 0.051$). The plot of the debiasing strategy by outcome interaction is depicted in Figure 2 (page 16). Due to the significant two-way interaction, the main effect of the outcome cannot be interpreted. Instead, simple main effect tests were conducted.

Simple main effect tests

In order to determine the effect of outcome information on auditor viability judgment, simple main effect tests consisting of a series of contrasts were conducted and are presented in Table 3. The means contrasted in Table 3 were taken from Table 1 and are the mean viability judgments made by the no outcome, failure outcome, and success outcome subjects in the no-debiasing-strategy condition. The no-debiasing-strategy mean viability judgments are summarized below:

No Outcome	60.66%
Failure Outcome	54.37%
Success Outcome	67.00%

In order to test H1, it is necessary to determine if the failure outcome mean viability judgment of 54.37% and the success outcome mean viability judgment of 67.00% are significantly different from the no outcome viability judgment of 60.66%. As revealed in Table 3, the failure outcome subjects' mean viability judgment of 54.37% is significantly less than the no outcome mean viability judgment of 60.66% ($p = .060$, one-tail probability). This indicates that, despite instructions to ignore the outcome information, being informed that the company failed caused the subjects in the failure outcome condition to judge continued viability as less likely than did the no outcome subjects. In other words, the failure outcome subjects were prone to hindsight bias.

In addition, Table 3 reveals that the success outcome subjects' mean viability judgment of 67.00% is significantly greater than the no outcome subjects' mean viability judgment of 60.66% ($p = 0.059$, one-tail probability). This indicates that, despite instructions to ignore the outcome information, being informed that the company continued caused the subjects in the success outcome condition to judge continued viability as more likely than did the no outcome subjects.

In short, both the failure outcome and the success outcome subjects in the no-debiasing-strategy condition were prone to hindsight bias. This provides support for H1; auditors with outcome information judged the reported outcome as more likely to occur than did auditors not provided with outcome information.

Table 3
Contrasts of Mean Viability Judgments Between Outcome Groups in the No-Debiasing-Strategy Condition

Outcome Group	No. of Subjects	DF	Mean	(SD)	Contrast	t
No Outcome	38		.6066	(.172)		
		74			.0629	1.55*
Failure Outcome	38		.5437	(.171)		
No Outcome	38		.6006	(.172)		
		74			.0634	-1.56*
Success Outcome	38		.6700	(.178)		

*Significant at the .05 Level, One-Tail Probability

Instructions to ignore outcome information

In order to find complete support for H1, it was necessary to determine whether the subjects denied using the outcome information. The subjects in the failure outcome and success outcome conditions were instructed to ignore the outcome information and to make their viability judgments as if they did not know the outcome. In a two-part debriefing question, the failure outcome and success outcome subjects were asked to indicate with a yes or no response whether or not they ignored the outcome information. If they answered no, they were then asked to indicate how much the outcome information influenced their viability judgment on a 7-point scale, from 1, slightly, to 7, considerably.

Among the 76 failure outcome subjects, 11 (14.5%) indicated that they did not ignore the outcome information, and the mean influence rating was 3.18. Among the 76 success outcome subjects, only 5 (6.5%) indicated that they did not ignore the outcome information, and the mean influence rating was 2.6.

The 2X3 ANOVA presented in Table 2 was performed again excluding the 16 subjects who indicated that they had not ignored the outcome information. Excluding these subjects did not significantly alter the ANOVA results. Given the relatively low number of subjects indicating that they did not ignore the outcome information combined with the low mean influence ratings, and given that these subjects' viability judgments did not significantly alter the results, it appears that the denial component of hindsight bias was achieved.

Results of hypothesis 2

H2 predicted that auditors informed of a failure outcome (i.e., an occurrence) would exhibit greater hindsight bias than would auditors informed of a success outcome (i.e., a nonoccurrence). As revealed in Table 3 and as previously discussed, both the failure outcome and success outcome subjects in the no-debiasing-strategy condition were prone to hindsight bias. The degree of hindsight bias exhibited by the failure outcome subjects can be measured by computing the absolute value of the difference between the failure outcome subjects' mean viability judgment (54.37%) and the no outcome subjects' mean viability judgment (60.66%) which is equal to 6.29%. Similarly, the degree of hindsight bias exhibited by the success outcome subjects can be measured by computing the absolute value of the difference between the success outcome subjects' mean viability judgment (67.00%) and the no outcome subjects' mean viability judgment (60.66%) which is equal to 6.34%.

In the no-debiasing-strategy condition, the degree of hindsight bias exhibited by the failure outcome subjects of 6.29% is not significantly different from the degree of hindsight bias exhibited by the success outcome subjects of 6.34% as revealed by the contrast in Table 4 (p=0.994, two-tail probability). As a result, H2 is not supported. However, in the debiasing strategy condition (to be discussed in more detail shortly), the failure outcome does cause greater hindsight bias as compared to the success outcome which provides partial support for H2.

Table 4
Contrasts of Mean Viability Judgment Differences Between the No Outcome and Failure Outcome Groups and the No Outcome and Success Outcome Groups in the No-Debiasing-Strategy Condition

Outcome Group	No. of Subjects	DF	Mean	(SD)	Difference In Means	Contrast	t
No Outcome	38	74	.6066	(.172)	.0629		
Failure Outcome	38		.5437	(.171)		.0005	.01
No Outcome	38	74	.6066	(.172)	.0634		
Success Outcome	38		.6700	(.178)			

Results of hypothesis 3

H3 predicted that when asked to self-generate a list of adverse factors and mitigating factors prior to the receipt of outcome information, auditors would record more adverse factors than mitigating factors and would rate the adverse factors as more relevant. The 114 subjects in the debiasing strategy group recorded an average of 6.58 adverse factors which, based on paired samples t-tests presented in Table 5, is significantly greater than the average number of mitigating factors recorded of 4.92 (p=0.002, one-tail probability). On a scale from 1, somewhat important, to 4, very important, the subjects assigned the adverse factors a mean importance rating of 3.21 which, based on paired samples t-tests presented in Table 6, is significantly higher than the mean importance rating of 2.91 assigned to the mitigating factors (p=0.000, one-tail probability). Thus, H3 is supported.

Table 5
Paired t-Test: Comparison Between Mean No. of Adverse Factors Recorded and Mean No. of Mitigating Factors Recorded

Factors	No. of Subjects	DF	Mean	(SD)	Difference	t
Adverse Factors	114	112	6.58	(2.069)	1.66	7.30*
Mitigating Factors			4.92	(1.938)		

*Significant at the .05 Level, One-Tail Probability

Table 6
Paired t-Test: Comparison Between Mean Importance Rating for Adverse Factors and Mean Importance Rating for Mitigating Factors

Factors	No. of Subjects	DF	Mean	(SD)	Difference	t
Adverse Factors	114	112	3.21	(.312)	30	6.79*
Mitigating Factors			2.91	(.467)		

*Significant at the .05 Level, One-Tail Probability

Results of hypothesis 4

H4a predicted that allowing auditors who have outcome information to review their lists of adverse factors and mitigating factors that they recorded in foresight would reduce hindsight bias for auditors provided with success outcome information. There is a significant interaction effect (p=0.051) in the 2X3 (debiasing strategy by outcome) ANOVA presented in Table 2. Therefore, to ascertain whether the debiasing strategy was successful, it was necessary to perform simple main effect tests. Contrasts of means were conducted to determine if the failure outcome and success outcome subjects in the debiasing strategy condition exhibited hindsight bias. The contrasts are present in Table 7.

Table 7
Contrasts of Mean Viability Judgments Between Outcome Groups in the Debiasing Strategy Condition

Outcome Group	No. of Subjects	DF	Mean	(SD)	Contrast	t
No Outcome	38	74	.6453	(.171)	.194	4.78*
Failure Outcome	38		.4513	(.198)		
No Outcome	38	74	.6453	(.171)	.040	-.99
Success Outcome	38		.6853	(.161)		

*Significant at the .05 Level, One-Tail Probability

Hindsight bias in debiasing strategy condition

The means contrasted in Table 7 were taken from Table 1 and are the mean viability judgments made by the no outcome, failure outcome, and success outcome subjects in the debiasing strategy condition. The debiasing strategy mean viability judgments are summarized below:

No Outcome	64.53%
Failure Outcome	45.13%
Success Outcome	68.53%

To determine whether the debiasing strategy subjects exhibited hindsight bias, the failure outcome mean viability judgment of 45.13% and the success outcome mean viability judgment of 68.53% were contrasted with the no outcome mean viability judgment of 64.53%. As revealed in Table 7, the failure outcome subjects' mean viability judgment of 45.13% is significantly less than the no outcome subjects' mean viability judgment of 64.53% (p=0.000, one-tail probability). This indicates that the failure outcome subjects in the debiasing strategy condition were prone to hindsight bias.

Table 7 also reveals that the success outcome subjects' mean viability judgment of 68.53% is not significantly greater than the no outcome subjects' mean viability judgment of 64.53% (p=0.163, one-tail probability). This indicates that the success outcome subjects in the debiasing strategy condition were not prone to hindsight bias. In short, both the failure outcome and success outcome subjects in the no-debiasing-strategy condition exhibited hindsight bias; however, only the failure outcome subjects exhibited hindsight bias in the debiasing strategy condition. Thus, the debiasing strategy was successful in eliminating hindsight bias for the success outcome subjects which supports H4a.

Debiasing strategy and failure outcome

As discussed, the failure outcome subjects exhibited hindsight bias in both the no-debiasing-strategy and the debiasing strategy conditions. As shown in Figure 2, the degree of hindsight bias exhibited by the failure outcome subjects increased from 6.29 (60.66 minus 54.37) in the no-debiasing-strategy condition to 19.4 (64.53 minus 45.13) in the debiasing strategy condition. As shown in Table 8, this increase is significant (p=0.012, one-tail probability). In addition, the contrast in Table 9 reveals that the failure outcome mean viability judgment in the no-debiasing-strategy condition of 54.37% is significantly greater than the failure outcome mean viability judgment in the debiasing strategy condition of 45.13% (p=0.012, one-tail probability). In short, the debiasing strategy was not successful in reducing hindsight bias for the failure outcome subjects, but instead increased the bias as predicted. Thus, H4b is supported.

Table 8
Contrasts of Mean Viability Judgment Differences Between the No Outcome and Failure Outcome Groups in the No-Debiasing-Strategy Condition and the Debiasing Strategy Condition

Outcome Group	No. of Subjects	DF	Mean (SD)	Difference In Means	Contrast	t
No Outcome-No-Debiasing-Strategy	38		.6066 (.172)			
		74		.0629		
Failure Outcome-No-Debiasing Strategy	38		.5437 (.171)			
					.1311	2.29*
No Outcome-Debiasing Strategy	38		.6453 (.171)			
		74		.1940		
Failure Outcome-Debiasing Strategy	38		.4513 (.198)			

*Significant at the .05 Level, One-Tail Probability

Table 9
Contrasts of Mean Viability Judgment Differences Between the No-Debiasing-Strategy Condition and the Debiasing Strategy Condition by Outcome Group

Outcome Group	No. of Subjects	DF	Mean	(SD)	Difference	t
No Outcome-No-Debiasing Strategy	38		.6066	(.172)		
		74			.0387	-.95
No-Outcome-Debiasing Strategy	38		.6453	(.171)		
Failure Outcome-No-Debiasing Strategy	38		.5437	(.171)		
		74			.0924	2.28*
Failure Outcome-Debiasing Strategy	38		.4513	(.198)		

*Significant at the .05 Level, One-Tail Probability

DISCUSSION AND CONCLUSION

The purpose of this study was to examine the effects of hindsight bias on auditors' going-concern judgments and the degree to which the bias is mitigated by a debiasing strategy found to be successful in the psychological literature. Consistent with prior auditing research (Anderson, 2000, 2002; Kennedy, 1993, 1995; Reimers & Butler, 1992), the current study found that auditors are prone to hindsight bias when making viability judgments (H1). H2 which predicted that hindsight bias is greater for the failure outcome as compared to the success outcome was not supported which is inconsistent with prior auditing research (Kennedy, 1995). However, it was found that a debiasing strategy by outcome interaction creates greater hindsight bias for the failure outcome in the debiasing strategy condition, thereby providing partial support for H2.

In addition, the study found that a preoutcome debiasing strategy found to be successful in the psychological literature (Davies, 1987) produces asymmetrical effects in an audit setting involving going-concern judgments. Due to the nature of auditor training coupled with the negative consequences of failing to correctly detect adverse factors, auditors self-generate, in foresight, a greater number of adverse factors and rate the adverse factors as more relevant as compared to mitigating factors (H3). Allowing auditors who have success outcome information to review their self-generated lists of adverse factors and mitigating factors eliminates hindsight bias (H4a). However, reviewing the lists of factors substantially increases hindsight bias in the failure outcome condition (H4b).

The presence of hindsight bias in the context of going-concern judgments may lead to a "knew-it-all-along" attitude, which impedes feedback learning (Fischhoff, 1975), thereby reducing what auditors could potentially learn from the feedback provided by actual bankruptcies. The presence of hindsight bias is particularly troubling in the case of the failure outcome. As compared to success outcomes, auditors have limited actual experience with failure outcomes. Also, because inaccurately predicting the failure outcome (i.e., issuing an unqualified opinion to a troubled company that subsequently fails) poses more dire consequences for public accounting firms than does inaccurately predicting the success outcome (i.e., issuing a modified opinion to a troubled company that continues), it is imperative that auditors learn as much as they can from troubled companies that fail. One could argue that Enron's auditors, the now defunct Arthur Andersen, certainly could have benefited from improved decision making regarding going-concern issues.

Due to the negative effects of hindsight bias, it is important to discover means by which it can be reduced. Given that monetary incentives (Camerer et al., 1989; Hell et al., 1988), accountability (Kennedy, 1993, 1995), and experience (Anderson, 2000; Kennedy, 1995) have been found ineffective in counteracting hindsight bias, it is important to discover decision aids that are successful in reducing it. In an auditing study involving an analytical review task, Kennedy (1995) found that a postoutcome debiasing strategy found to be successful in the psychological literature (Davies, 1987; Slovic & Fischhoff, 1977) is successful in an auditing context. Although Kennedy (1995) has already found that a postoutcome debiasing strategy is effective in eliminating hindsight bias in an audit setting, it is important to also examine the effectiveness of a preoutcome debiasing strategy for two main reasons. First, although Davies (1987) found preoutcome and postoutcome debiasing strategies to be equally effective, he also points out that for events whose occurrence and importance are known in advance (such as elections, space launches, and impending bankruptcies), a preoutcome debiasing strategy may be more useful. Second, given that a preoutcome debiasing strategy may be more useful in certain situations, it is important to discover to what extent it is indeed effective in an audit setting. As the current study has shown, Davies' preoutcome debiasing strategy eliminates hindsight bias in the success outcome condition, but increases the bias in the failure outcome condition.

The main contribution of this study is that debiasing strategies found to be successful in reducing hindsight bias in the psychological literature may produce asymmetrical effects in an auditing environment by eliminating the bias for some outcomes and exacerbating it for others. Due to auditors' unique training and experience, it cannot be assumed that auditors will behave and respond in the same manner as do subjects in psychological experiments. This illustrates the need to exercise caution when importing results from the psychological literature to an auditing domain. It may first be necessary to subject the findings in the psychological literature to empirical testing that includes auditor subjects performing auditing tasks.

The current study's results also raise the question as to whether or not Kennedy's (1995) postoutcome debiasing strategy would be effective in a going-concern context. Using an analytical review task, Kennedy instructed her subjects to generate reasons as to why the outcome (either high or low 13^{th} quarter's sales for two products) might be considered unlikely. There is no basis to expect subjects to generate significantly more reasons for the high outcome than the low outcome or vice versa, and Kennedy's results do not indicate that either the high or low outcome elicited significantly more reasons. However, using a going-concern task, auditors provided with the failure outcome may be unable to generate many mitigating factors in hindsight. This would cause the failure outcome to appear even more likely, thereby increasing hindsight bias. Auditors provided with the success outcome should be able to generate a greater number of adverse factors which would make the success outcome appear less likely, thereby reducing hindsight bias.

The results of the study must be interpreted in light of certain limitations. First, the study involves a sample of auditor subjects from international public accounting firms which limits the ability to generalize the results to smaller public accounting firms at the national, regional, and local levels. Second, it is difficult to determine whether the subjects were sufficiently motivated to concentrate on the experimental tasks and to complete the tasks as they would in practice. Third, the subjects did not have access to the array of information, resources, and consultations with others that would normally be available to them during an actual audit. Also, the subjects may not have been able to relate to many situations in practice in which they are required to ignore known outcomes and state explicitly what judgments they would have made at some point in the past.

Future research could further examine when it is best to debias the effects of hindsight using a preoutcome strategy versus a postoutcome strategy. In many situations, preoutcome strategies may be impractical because, as Davies (1987) points out, the event's importance may not be evident in foresight, or it may not even be known that an event is going to occur. For those events whose importance and occurrence are known in advance, preoutcome strategies may be more useful. However, as the current study reveals, preoutcome strategies must be used with caution and appear to be more useful when a fairly equal number of reasons for the alternative outcomes can be generated. Future research could also further investigate a debiasing strategy that is successful in reducing hindsight bias when auditors are informed that a troubled company failed. Perhaps providing auditors with lists of mitigating factors rather than instructing them to self-generate the factors would successfully reduce the bias.

REFERENCES

Ahrens, A. A., Elder, R. J., & Beasley, M. S. (2006). *Auditing and assurance services: An integrated approach.* Upper Saddle River, NJ: Prentice Hall.

American Institute of Certified Public Accountants. (1988). Statement on Auditing Standards No. 59. *The Auditor's Consideration of an Entity's Ability to Continue as a Going Concern.* New York, NY: AICPA.

_____. (1990). Statement on Auditing Standards No. 64. *Omnibus Statement on Auditing Standards – 1990: The Auditor's Consideration of an Entity's Ability to Continue as a Going Concern.* New York, NY: AICPA.

Anderson, J. C., Jennings, M. M., Kaplan, S. E., & Reckers, P. M. J. (1995). The effect of using diagnostic decision aids for analytical procedures on judges' liability judgments. *Journal of Accounting and Public Policy, 14*(1), 33-62.

Anderson, J. C., Jennings, M. M., Lowe, D. J., & Reckers, P. M. J. (1997). The mitigation of hindsight bias in judges' evaluations of auditor decisions. *Auditing: A Journal of Practice & Theory, 16*(2), 20-39.

Anderson, K. L. (2000). Hindsight bias and auditors' going-concern judgments: The effects of experience. *Pennsylvania Journal of Business & Economics, 7*(1), 1-18.

_____. (2002). How outcome information in hindsight influences auditors' relevance ratings of adverse factors versus mitigating factors. *Journal of Business & Information Technology, 2*(1), 1-13.

Arkes, H. R., Faust, D., Guilmette, T. J., & Hart, K. (1988). Eliminating the hindsight bias. *Journal of Applied Psychology, 73*(2), 305-307.

Asare, S. (1992). The auditor's going-concern decision: Interaction of task variables and the sequential processing of evidence. *The Accounting Review, 67*(2), 379-393.

BankruptcyData.com. (2006). *The 15 largest bankruptcies 1980 – present.* Retrieved June 21, 2006, from http://www/bankruptcydata.com/

Biggs, S. F., & Wild, J. J. (1985). An investigation of auditor judgment in analytical review. *The Accounting Review, 60*(3), 607-633.

Brown, C. E., & Solomon, I. (1987). Effects of outcome information on evaluations of managerial decisions. *The Accounting Review, 62*(3), 564-577.

Buchman, T. A. (1985). An effect of hindsight on predicting bankruptcy with accounting information. *Accounting, Organizations, and Society, 10*(3), 267-285.

Camerer, C., Loewenstein, G. F., & Weber, M. (1989). The curse of knowledge in market settings: An experimental analysis. *Journal of Political Economy, 97*(5), 1232-1254.

Davies, M. (1987). Reduction of hindsight bias by restoration of foresight perspective: Effectiveness of foresight-encoding and hindsight-retrievability strategies. *Organizational Behavior and Human Decision Processes, 40*(2), 50-68.

Fischhoff, B. (1975). Hindsight = foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance, 1*(3), 288-299.

__________. (1977). Perceived informativeness of facts. *Journal of Experimental Psychology: Human Perception and Performance, 3*(2), 349-358.

__________. (1982). For those condemned to study the past: Heuristics and biases in hindsight. In D. Kahneman, P. Slovic, and A. Tversky (Eds.), *Judgment Under Uncertainty: Heuristics and Biases* (pp. 335-351). New York: Cambridge University Press,.

Fischhoff, B., & Beyth, R. (1975). I knew it would happen -remembered probabilities of once-future things. *Organizational Behavior and Human Performance, 13*(1), 1-16.

Hasher, L., Attig, M. S., & Alba, J. W. (1981). I knew it all along: Or did I? *Journal of Verbal Learning and Verbal Behavior, 20*(1), 86-96.

Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin, 107*(3), 311-327.

Heintz, J. A., & White, G. B. (1989). Auditor judgment in analytical review -- some further evidence. *Auditing: A Journal of Practice and Theory, 8*(1), 22-39.

Hell, W., Gigerenzer, G., Gauggel, S., Mall, M., & Mueller, M. (1988). Hindsight bias: An interaction of automatic and motivational factors? *Memory and Cognition, 16,* 533-538.

Kennedy, J. (1993). Debiasing audit judgment with accountability: A framework and experimental results. *Journal of Accounting Research, 31*(2), 231-245.

_____. (1995). Debiasing the curse of knowledge in audit judgment. *The Accounting Review, 70*(2), 249-273.

Kida, T. E. (1984). The impact of hypothesis-testing strategies on auditors' use of judgment data. *Journal of Accounting Research, 22*(2), 332-340.

Kinney, W. R., & Uecker, W. C. (1982). Mitigating the consequences of anchoring in auditor judgments. *The Accounting Review, 57*(1), 55-69.

Loftus, E. F., & Loftus, G. R. (1980). On the permanence of stored information in the human brain. *American Psychologist, 35*(5), 409-420.

Lowe, D. J., & Reckers, P. M. J. (1994). The effects of hindsight bias on jurors' evaluations of auditor decisions. *Decision Sciences, 25*(3), 401-426.

Maddocks, P. M. (1989). Outcome knowledge and auditor judgment. Ph.D. Dissertation. University of Florida.

Mazursky, D., & Ofir, C. (1990). "I could never have expected it to happen": The reversal of the hindsight bias. *Organizational Behavior and Human Decision Processes, 46*(1), 20-33.

McDaniel, L. S., & Kinney, W. R., Jr. (1994). Expectation-formation guidance in the auditor's review of interim financial information. *Journal of Accounting Research, 32*(2), 275-291.

Reimers, J. L., & Butler, S. A. (1992). The effect of outcome knowledge on auditors' judgmental evaluations. *Accounting, Organizations and Society, 17*(2), 185-194.

Schkade, D. A. & Kilbourne, L. M. (1991). Expectation-outcome consistency and hindsight bias.

Organizational Behavior and Human Decision Processes, 49(1), 105-123.

Slovic, P., & Fischhoff, B. (1977). On the psychology of experimental surprises. *Journal of ExperimentalPsychology: Human Perception and Performance*, *3*(4), 544-551.

Trotman, K., & Sng, J. (1989). The effect of hypothesis framing, prior expectations, and cue diagnosticity on auditor's information choice. *Accounting, Organizations, and Society*, *14*(5-6), 565-576.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124-1131.

Wasserman, D., Lempert, R.O., & Hastie, R. (1991). Hindsight causality. *Personality and Social Psychology Bulletin, 17*(1), 30-35.

Wood, G. (1978). Knew-it-all-along effect. *Experimental Psychology: Human Perception and Performance, 4*(2), 345-353.

Kim L. Anderson, University of Pittsburgh, is an associate professor at Indiana University of Pennsylvania. Her research interests include behavioral research in accounting, particularly the effects of various biases on auditors' judgments.

ENDNOTES

1. In addition, auditors who issue unqualified opinions to client companies that subsequently fail may be unfairly evaluated, in hindsight, by interested third parties such as the SEC, stockholders, expert witnesses, jurors, and peers (Kennedy, 1995; Lowe & Reckers, 1994). These third parties may be unable to ignore the outcome information they have (i.e., the company did indeed fail) that the auditors did not have at the time they made their opinion decision.

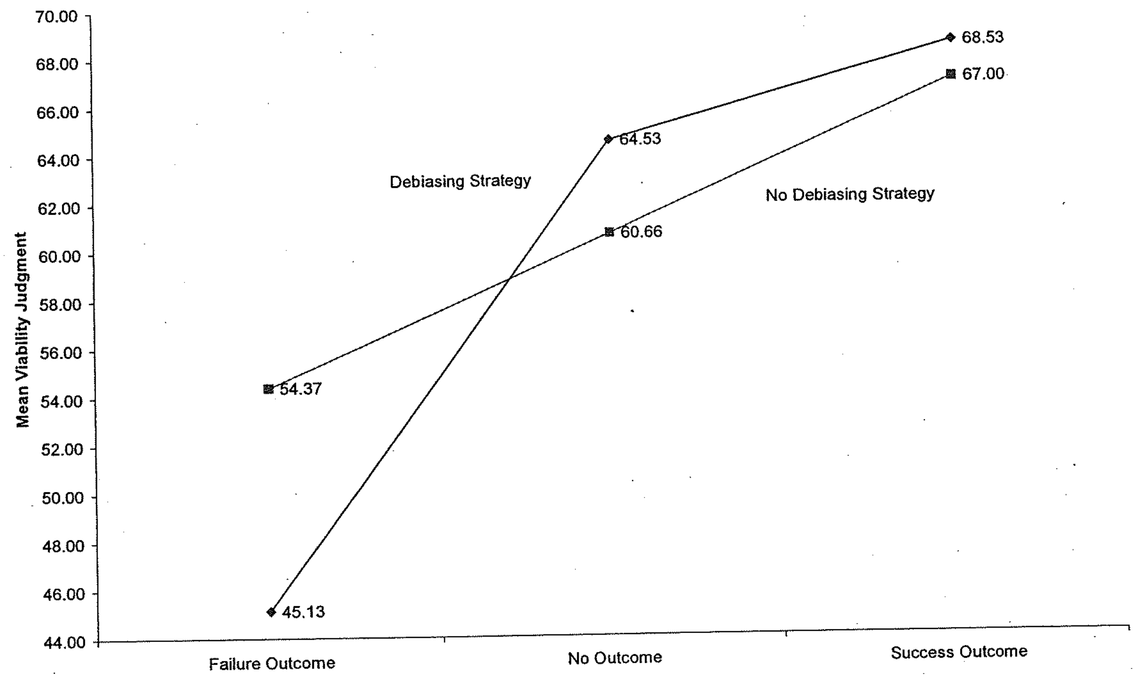
2. Creeping determinism is also consistent with Loftus and Loftus's (1980) suggestion that memory for complex events will be erased and updated by new information when it is inefficient or inconsistent to maintain two different memories. According to this view, the foresight state of mind cannot be recaptured in hindsight (e.g., Mazursky & Ofir, 1990). When instructed to ignore outcome information, individuals unable to recapture their foresight perspectives, may use Tversky and Kahneman's (1974) representativeness or availability heuristics to make their judgments (Fischhoff, 1975).

3. The positive and negative factors included in this study were selected from a list of adverse and mitigating factors originally used in Kida (1984) and subsequently used in other going-concern studies (e.g., Trotman & Sng, 1989). The adverse factors included: "Management indicated that new legislation may make it difficult for Alpha (the case company) to market one of their major products," "Discussions with management indicate that a material liability from litigation is likely this year," "Management indicates that there is a good chance of losing a major customer," and "Management and labor representatives indicate that there is a chance that labor will strike this year."

The mitigating factors included: "Alpha's technology is competitive with other firms in the chemical industry," "The economic outlook for the chemical industry is stable," "In general, Alpha's suppliers indicate that usual trade credit to Alpha will be available," and "Management states that it is possible that a key patent may be obtained in the near future." The sample case was adapted from one used by Maddocks (1989) and was selected because it was a company experiencing moderate financial problems, but not necessarily on the brink of failure. It was important to select a case that was not obviously financially sound or on the verge of bankruptcy because the same case was used for both the Success and Failure Outcome conditions.

4. Previous studies have manipulated the framing of this question to see whether auditors responded differently to being asked how likely the company was to succeed versus how likely it was to fail (e.g., Kida, 1984; Trotman & Sng, 1989; Asare, 1992). Auditors listed more failure reasons than success reasons and rated failure reasons as more relevant than success reasons in both frames, indicating that the conservatism effect dominates any framing effect. In this study, we use only one frame, asking subjects to judge the likelihood of the company continuing because this is the way the task is framed on an actual audit.

FIGURE 2 Debiasing Strategy by Outcome Interaction



DEVELOPING A PRACTICAL FORMULA FOR OPTIMAL CAPITAL STRUCTURE

Joseph Cheng, Ithaca College
Alka Bramhandkar, Ithaca College

ABSTRACT

The model developed in this paper may be considered a breakthrough in the area of corporate finance in that it is the first of its kind to apply the theory of optimal capital structure to the real world. While the theory of capital structure is well discussed in the literature, the theory has never been brought down to an application level. The model in this paper enables corporate managers to actually calculate the optimal debt ratio for their corporations.

INTRODUCTION

One of the most important decisions that management must make relates to how much a company should borrow. The tax advantage associated with issuing debt would suggest that companies should borrow at high levels. However, most companies monitor their debt levels to maintain their credit rating and minimize the effect of bankruptcy costs. General economic conditions, management conservatism, a company's asset structure and expected growth rate are some of the important considerations in deciding the optimum capital structure.

This paper is organized as follows: 1. a brief summary of the literature review on capital structure. 2. the conditions for optimal capital structure which would lead to share price maximization. 3. the results are tested on 23 companies in the utility industry. 4. the conclusion summarizes the advantages of this approach.

LITERATURE REVIEW

Several ideas have been proposed to explain the capital structure decision. These can be classified in five main groups:

Effect of taxes

In their modified paper, Modigliani and Miller 1963) suggested that shareholder wealth can be increased by borrowing because the interest on debt is tax deductible, making debt a cheaper source of financing.

Trade-off theory

These theories suggested that the tax advantage associated with debt is offset at higher levels of debt by bankruptcy-related costs forcing managers to borrow less (DeAngelo & Masulis, 1980) (Leland, 1994).

Pecking order

In 1984, Steward Myers offered a completely different explanation stating that firms have a strong preference for internal financing as their first choice. External debt comes next, with equity financing used only as a last resort. Equity financing leads to ownership dilution and carries higher costs. Ghosh (1999) discusses and tests a modification to this theory. His results indicate that firms converge to the industry mean while following the idea of the pecking order.

Signaling

Several recent models have applied the idea of using debt as a signal to convey information to the capital market. This theory implies that issuing debt is an indication that the firm has the ability to service it, whereas stock issues can be construed as a negative event, suggesting that the stock is overpriced (Myers & Majluf, 1984). Lucas and McDonald (1990) and Fama (1985) show positive stock price response to firm announcements of bank debt agreements, suggesting that the loan approval was based on access to favorable private information.

Target

Bowen, Daley, and Huber, Jr. (1982) presented a hypothesis that companies attempt to achieve the average industry ratio over time. Their conclusion was supported by the data over five- and ten-year time periods. Similar conclusions are also reached by Marsh (1982) and Jalilvand and Harris (1984). Claggett (1991) found that although unusual conditions may prevent companies from adjusting towards the industry mean, convergence to the mean is more common for companies with higher than the industry debt levels. Hull (1999) investigates the stock price response for two groups—firms moving away from the industry benchmark and firms moving closer to the industry benchmark. He finds significantly more negative returns for the first group.

Although there is a plethora of literature regarding the theory of capital structure, application of the theory is rather limited. Most models require empirical estimation of certain parameters or require the use of certain data that are not readily observable. However, empirical estimation involves exposure to many problems such as specification errors, data availability, and measurement errors. While it is well known that optimal capital structure is attained when the weighted average cost of capital is at a minimum, there is a lack of formulas useful to corporate financial managers for actually computing the optimal capital mix for their firms. Gitman (2003) used a simulation approach for calculating share prices at various levels of debt. Taking this to a higher level, we have developed a formula from share price maximization that enables financial mangers to generate the optimal debt level directly without resorting to simulations or statistical analysis. Furthermore, the data needed for applying this model are readily obtainable.

In the following sections, we present a model to derive the conditions for optimal capital structure, followed by a discussion of how the Capital Asset Pricing Model (CAPM) is utilized to derive the values for certain derivatives, so that the optimal debt level can be estimated.

THE FORMULA FOR OPTIMAL CAPITAL STRUCTURE

The objective of this paper is to derive quantitatively the capital structure (debt ratio) for firms which maximizes share price in the context of the Capital Asset Pricing Model (CAPM). It has been shown that share price maximization is consistent with minimization of weighted average cost of capital.

Before share price is maximized with respect to debt level, the share price equation must first be specified. For firms with no growth, share price in the standard constant growth model is written as:

$$P = \frac{EPS}{R_e} \tag{3}$$

where R_e= required return for stock

The value in equation (3) can be expressed alternatively in terms of earnings of the entire firm instead of earnings per share. This means that the market value of equity can be expressed as profit after taxes divided by R_e:

$$P\,S = \frac{[EBIT - R_d(D_o + d)]n}{R_e} \tag{4}$$

where

EBIT= earnings before interest and taxes

R_d= interests rate for debt

P = price after asset mix adjustment

S = number of shares outstanding after mix adjustment

D_o = beginning balance of debt

d = new debt to be issued

n = 1-tax rate (net after taxes)

Since this paper focuses on the optimal mix of capital rather than the optimal level of capital, the amount of total capital is to be held constant throughout the price maximization process. It is important to keep the amount of capital investment separated from the financing aspects. Changing the capital mix without changing the total amount of capital means that any issuance of new debt must be offset by repurchase of stocks in equal amount or vice versa.

$$d = P(S_o - S) \tag{5}$$

where d denotes the amount of new debt to be issued

S_o= original number of shares outstanding

The term (S_o-S) in (5) represents the change in shares outstanding. Note that, if S_o is greater than S, then d is greater than 0, which means that the amount of debt to be issued equals the value of stocks to be repurchased. And if S_o is less than S, then d is less than 0, which means the value of shares to be issued equals the amount of debt to be retired. Thus, the total amount of capital is neither increased nor decreased by the issuance of new debt. Any change in debt in this case is merely an adjustment to the capital mix. Thus, (5) will be referred to as the constant capital constraint, which ensures that the level of total investment remains unchanged throughout the adjustment process. Since the market is efficient, the share price at which the shares are issued or repurchased (P) should have already reflected the capital mix decision.

The constant capital constraint (5) can be rearranged so that it becomes:

$$PS = PS_0 - d \tag{6}$$

which can be imposed on (4) by substituting it into the left hand side of (4):

$$PS_0 = \frac{[EBIT - R_d(D_0 + d)]n + R_e d}{R_e} \tag{7}$$

P So in (7) represents the value of equity for the current or incumbent shareholders after the capital mix adjustment. Since the number of existing shares outstanding (S_o) is a constant, maximizing P S_o is equivalent to maximizing share price P. And since the constant capital constraint has already been imposed, maximizing (7) with respect to d would generate the optimal debt ratio that maximizes share price. And this is what we seek to derive. Before differentiating (7) with respect to d, we need to assess which variables in (7) would be affected by the change in d. In this case, R_d and R_e would rise as d increases because of greater financial risk to be added by having more debt. Thus, the derivative of (7) with respect to d for finding the optimal capital mix is:

$$(PS_0)' = R_e\left(-R_d' D_0 n - R_d' dn - R_d n + R_e' d + R_e\right) \tag{8a}$$

$$-R_e'\left(EBIT n - R_d D_0 n - R_d dn + R_e d\right) = 0$$

where the apostrophes in (8a) denote partial derivatives with respect to debt issued (d).

For example, $R_e' = \dfrac{\partial(R_e)}{\partial(d)}$

Dividing both sides of (8a) by R_e and rearranging terms, we get

$$-R_d' D_0 n - R_d' dn - R_d n + R_e' d + R_e = R_e' \frac{(EBIT n - R_d D_0 n - R_d dn + R_e d)}{R_e} \tag{8b}$$

Using (7) on the right hand side of (8b) and rearranging terms, we arrive at the optimality condition:

$$R_e = R_d n + R_d'(D_0 + d)n + R_e'(PS_0 - d) \tag{9}$$

The term (D_o +d) in (9) represents original debt plus new debt, which combine to become D (new level of debt). Based on the constant capital constraint (5), the term (P S_o-d) in (9) is equivalent to P S, the market value of equity after the mix adjustment, which can be represented by E:

$$E = PS = PS_0 - d \tag{10}$$

(10) can be substituted into the last term in (9). Equation (9) can now be expressed as

$$R_e = R_d n + R_d' D n + R_e' E \tag{11}$$

which is the optimality or first-order condition derived by maximizing share price subject to the constant capital constraint. The second-order condition showing that share price is maximized is presented in the Appendix.

Let us analyze the economic implications of the optimality condition stated in (11). The sum of the three terms on the right hand side of (11) represents the total marginal cost of debt: the first term represents the after tax cost of debt for marginal or incremental debt; the second term represents the incremental after-tax costs on inframarginal or existing debt (as d rises, the required return on debt (R_d) increases for all debts); such rise in interest cost for inframarginal debts due to higher required yield is another cost of debt which should be considered; the third term represents the increase in the cost of equity caused by heavier debt, which causes earnings to be more volatile and thus creates greater risk to shareholders. This last term is considered a cost associated with new debt although it is channeled through the cost of equity.

The sum of the after-tax interest costs on the incremental debt, the increase in interest cost on the inframarginal or existing debt, and the cost of higher required return for equity as a result of having more debt, constitute the total marginal cost of debt. Thus, the optimality condition (11) states that debt should be utilized to the extent where total marginal cost of debt equals the cost of equity which the new debt replaces. The formula for optimal debt will be derived from (11) in the following sections.

USING THE CAPITAL ASSET PRICING MODEL (CAPM)

In order to arrive at the optimal debt in actual numerical terms using (11), we need to first compute R_e' and R_d', which may be interpreted as the marginal cost of debt channeled through the required return on equity and the marginal cost of debt channeled through interest rate, respectively. While the values for these two partial derivatives are unobservable, they may be derived from the Capital Asset Pricing Model as follows.

According to the Capital Asset Pricing Model, the required return for investment is a function of beta. Thus, debt affects the required return through its impact on beta (B):

$$Re' = \frac{\partial Re}{\partial B}\frac{\partial B}{\partial d} \qquad (12)$$

Based on the standard CAPM equation, the first term in (12) is simply the market risk premium. Thus, (12) can be written as

$$Re' = [E(r_m) - r_f]\frac{\partial B}{\partial d} \qquad (13)$$

Based on (12) and (13), $R_{e'}$ can be computed only after the effect of debt on beta has been estimated. However, the effect of debt on beta is not a clear cut issue—many empirical studies have been conducted to estimate the effect of debt on beta, but different results have been observed. In this paper, the effect of debt on beta is not estimated empirically. Instead, it is derived by utilizing a debt-beta relationship established by Hamada (1978):

$$B = B_u [1+(1-t) D/E] \qquad (14)$$

where B= beta value for the firm

B_u= beta of the firm if unlevered

NUMERICAL SIMULATION

For the numerical simulation below, we generate the value of incumbent equity (PS_o) at various levels of debt for a hypothetical firm. We employ rates and values for a typical firm as follows:

Asset =100
EBIT =20
R_m =.12
R_f = .05
Tax Rate = .4
B_u = 1
R_d = 2/3 R_e

The last equation is based on the assumption that the required return for bond is 2/3 of that for stock. The simulation begins at zero debt, and then increases debt incrementally by 10 (as equity is being reduced by the same amount to maintain the same level of asset) for each interval. The level of debt is thus raised until it reaches 90 or 90% because a firm cannot have 100% debt and must have at least some equity. In total, there are ten levels of debt in this simulation, ranging from 0% to 90%.

For each level of debt, incumbent shareholders' value based on (7) and the first derivatives based on (8a) are generated. These simulated values are presented in Table 1.

Table 1
Equity Value at Various Debt Levels

D	First Derivative	PSo
0.000000	0.3600000E-02	100.0000
10.00000	0.3102844E-02	102.2567
20.00000	0.2343150E-02	103.9540
30.00000	0.1140686E-02	104.9565
40.00000	-0.8576000E-03	105.0811
50.00000	-0.4413600E-02	104.0741
60.00000	-0.1140660E-01	101.5738
70.00000	-0.2748560E-01	97.04587
80.00000	-0.7623360E-01	89.66667
90.00000	-0.3551976	78.09639

As seen above, the first derivative decreases as debt increases and reaches near zero at about 40% debt. (To be precise, the first derivative reaches zero at 36%, which is the optimal level of debt in this example.) This is also where the incumbent shareholders' value (PS_o) reaches its maximum level. The above numerical simulation demonstrates that the equity value for the incumbent shareholders indeed reaches a maximum when the first order condition is zero. In the next section, we derive the formula for directly calculating the optimal level of debt for a firm without a simulation.

DERIVING OPTIMAL DEBT FORMULA

Substituting D_o + d for D in (14), we can write

$$B= B_u [1+(1-t) (D_o +d)/E] \qquad (15)$$

Equation (15) illustrates the relationship between the level of debt and beta (B) as implied by the Capital Asset Pricing Model. The equation indicates that the value of beta increases with debt. Such correlation can be derived by differentiating (15) with respect to d. Before differentiation, the constant capital constraint should be imposed first by substituting (10) for E in (15), which becomes

$$B = B_u\,[1+(1\text{-}t)\,(D_o+d)/(P\,S_o\text{-}d)] \qquad (16)$$

Since $\dfrac{\partial(P\,S_o)}{\partial d} = 0$ at the optimal debt level, the partial derivative of (16) is simply:

$$\frac{\partial B}{\partial d} = \frac{B_u(1-t)\,[(P\,S_o - d) - (D_o + d)(-1)]}{(P\,S_o - d)^2} \qquad (17)$$

Substituting (10) for ($P\,S_o - d$) and D for (D_o +d) in (17), we can simplify (17) to:

$$\frac{\partial B}{\partial d} = \frac{B_u(1-t)(D+E)}{E^2} \qquad (18)$$

We now have derived the term $\dfrac{\partial B}{\partial d}$, which is one of the components determining R_e' as stated in (13).

Equation (18) captures the impact of debt on beta, which in turn affects the required return on equity (R_e). Substituting (18) into (13) would reveal in details the partials of R_e with respect to debt (R_e'), which is needed in the optimality condition as stated in (11).

In this paper, we assume that the required return for debt or yield for bonds is proportionate to the required return for equity. This means that the change in R_d is also proportionate to the change in R_e:

$$R'_d = k\,\frac{\partial R_e}{\partial B}\,\frac{\partial B}{\partial d} \qquad (19)$$

where $k = R_d/R_e$ (20)

In order to modify the optimal condition as stated in (11) into a form where optimal debt can be solved, we enter the above partial derivatives into the optimality condition (11), which becomes

$$Re = R_d\,n + \frac{\partial R_d}{\partial d}\,Dn + \frac{\partial R_e}{\partial B}\,\frac{\partial B}{\partial d}\,E \qquad (21)$$

Using (18), (19), and (20) in (21), we get

$$Re = R_d n + k\,\frac{\partial R_e}{\partial B}\,B_u(1-t)\frac{(D+E)}{E^2}\,Dn + \frac{\partial Re}{\partial B}\,B_u(1-t)\frac{(D+E)}{E^2}\,E \qquad (22)$$

Since n= (1-t), (22) can be simplified into

$$Re = R_d n + k\,\frac{\partial R_e}{\partial B}\,B_u\,n\left(\frac{D^2}{E^2} + \frac{D}{E}\right)n + \frac{\partial Re}{\partial B}\,B_u\,n\left(\frac{D}{E}+1\right) \qquad (23)$$

Since (23) is a detailed version of the optimality condition, the value of D/E that satisfies the condition as stated in (23) is the optimal debt to equity ratio. To solve for the optimal ratio, we rearrange the terms in (23) so that the optimality condition becomes a second-degree polynomial function of D/E:

$$a\,(D/E)^2 + b\,(D/E) + c = 0 \qquad (24)$$

$$a = k\,\frac{\partial R_e}{\partial B}\,B_u\,n^2$$

$$b = k\,\frac{\partial R_e}{\partial B}\,B_u\,n^2 + \frac{\partial R_e}{\partial B}\,B_u\,n$$

where

$$c = -Re + R_d\,n + \frac{\partial Re}{\partial B}\,B_u\,n$$

$$*D/E = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \qquad (25)$$

We have successfully developed a practical formula (25) which financial managers can actually use to derive the optimal debt ratio for their firms. This is indeed a breakthrough.

To apply the formula, actual financial data are used for generating the numerical values for a, b, and c. Once the numerical values for a, b, and c are calculated, the optimal debt ratio can be computed using (25). The source of data and computation procedures are described in the following section.

APPLICATION OF THE MODEL

To demonstrate how the optimal debt ratio can be calculated with real world data, we chose the utility industry as a proxy for the zero growth industry. Initially, all companies included in the "utility" industry by Value Line were analyzed to determine if relevant data needed for the application of the proposed model were available. All financial information is based on reported results for the year 2001-2003. The sample size was originally 26. Due to incomplete data availability for three firms, the final sample size is reduced to 23. A summary of the source of the data and how they are calculated is presented below.

Beta: as reported by Value Line. B_u is derived from the actual value of beta using (15).

$$\text{Tax rate} = \frac{\text{Total tax liability for the year}}{\text{Taxable income}}$$

Debt: long-term debt and preferred stock book values.

Equity market value: based on the closing stock price as of the end of each year, multiplied by the number of shares outstanding. Since equity is measured by market value, the debt to equity ratio in this paper is essentially market value based.

R_e: Required rate of return on equity as computed by the Capital Asset Pricing Model. The risk free rate is assumed to be .05 and the market premium .07.

R_d: Yield to maturity (at year-end) for corporate bonds obtained from the Mergent Bond Record.

Table 2 (page 27) presents the actual debt ratios along side with the optimal debt ratios as computed by the model.

The above demonstrates how the optimal levels of debt are calculated for the utility industry. With the newly available data for optimal debt, it is now possible to conduct empirical research on capital structure. Such research was difficult because there has been a lack of data for optimal debt. To demonstrate how the model developed in this paper can be applied to financial research relating to debt, we examine the relationship between managers' aversion to risk and their decision to take on less debt than the optimal level suggested by our model. It has been hypothesized that managers are more concerned about their own job security than maximizing stockholder wealth. Managers working in firms with a higher level of business risk, as measured by the unlevered beta, might keep the debt level below the optimum to minimize the possibility of bankruptcy. This hypothesis has never been tested because no one has actually computed the optimal debt level.

We use DIF, the difference between the actual and optimal debt levels, to measure the degree of debt overutilization. A positive value for DIF indicates that the firm's debt level exceeds the optimal level needed to maximize the stock price. On the other hand, negative DIF points to underutilization of debt.

The numerical results are presented in Table 2. Note that more firms (19 firms) underutilized debt in 2001 than in the other two years. This may be due to a combination of a gloomy economy in 2001 and the unexpected event of September 11 that rendered firms more pessimistic about the future than justified, causing them to be highly conservative and thus to take on less debt than optimal.

To see whether there is any correlation between the degree of under-utilization of debt and business risk, we regress DIF against the unlevered beta of the firms. Our independent variable is unlevered beta (B_u), which is a proxy for the level of business risk associated with the firm. Unlevered beta takes out the financial risk that has been incorporated in the levered beta. It is obtained by solving for B_u alone in (14):

$$B_u = \frac{B}{1 + (1-t)\frac{D}{E}} \qquad (26)$$

We regress

$$\text{DIF} = a + b\,(\text{unlevered beta}) \qquad (27)$$

Our regression results are reported in Table 3.

Table 3
Regression Summary

	Year 2001	Year 2002	Year 2003
Estimated Coefficient (b)	-2.59	-1.98	-.80
t-Statistic	(4.15)	(3.48)	(2.11)
R-Square	.45	.37	.17

Since the estimated coefficient values are all negative and statistically significant at the 95% level, the regression results support the hypothesis that firms with high unlevered betas (higher level of business risk) tend to have lower values of DIF (i.e., underutilization of debt.)[1] It is clearly prudent for managers of firms with higher business risk to be conservative with their use of debt. However, our study suggests that they might be overly conservative, perhaps putting their job security ahead of wealth maximization.

CONCLUSION

Based on a small sample of firms in the utility industry, we have shown that it is possible to calculate the optimal debt level for a mature or low growth industry. No assumption regarding the probability distribution on the earnings needs to be made for estimating the impact of debt on earnings or returns. In addition, the data needed to apply our model are readily available. Since our model for deriving optimal debt ratio does not require statistical analysis or regression, many potential problems and errors associated with these methods are avoided.

The model in this paper generates a new data set (optimal debt) for researchers to work with. This data set could open a new path for future statistical studies regarding capital structure. As a demonstration, we have conducted a simple statistical study utilizing such data. The results lend support to the premise that managers operating in the firms with a higher level of business risk tend to be overly conservative in their use of debt. By operating at a debt level below the optimum, they are not maximizing share prices and thus are not working in the best interest of the shareholders.

In the future, we plan to generalize the model so it can be applied to industries with positive growth. Our long-term plan is to derive a model without the assumption of zero or constant growth so that the model may be applied in a totally general way. Furthermore, instead of using the proportionality assumption for required return on debt and required return on equity, we seek to derive the required return on debt from a model. Ultimately, a complete model would be one that incorporates the issue of dividend into the capital structure problem so that both optimal dividend and optimal debt may be determined simultaneously.

REFERENCES

Bowen, R. M., Daley, L. A., & Huber, C. C., Jr. (1982). Evidence on the existence and determinants of inter-industry differences in leverage. *Financial Management*, *11*(4), 10-20.

Claggett, E. T., Jr. (1991). Capital structure: Convergent and pecking order evidence. *Review of Financial Economics*, *1*(1), 35-48.

DeAngelo, H., & Masulis, R. (1980). Optimal capital structure under corporate and personal taxation. *Journal of Financial Economics*, *8*(1), 3-29.

Fama, E. (1985). What's different about banks? *Journal of Monetary Economics*, *15*(1), 29-39.

Gitman, L. (2003). *Principles of managerial finance*. Boston, Massachusetts: Addison Wesley.

Ghosh, A., & Cai, F. (1999). Capital structure: New evidence of optimality and pecking order theory. *American Business Review*, *17*(1), 32-38.

Hull, R. M. (1999). Leverage ratios, industry norms and stock price reaction: An empirical investigation of stock-for-debt transactions. *Financial Management*, *28*(2), 32-45.

Jalilvand, A., & Harris, R. S. (1984). Corporate behavior in adjusting to capital structure and dividend targets: An econometric study. *Journal of Finance*, *39*(1), 127-145.

Leland, H. E. (1994). Corporate debt value, bond covenants, and optimal capital structure. *Journal of Finance*, *49*(4), 1213-1252.

Lucas, D., & McDonald, R. (1990). Equity issues and stock price dynamics. *Journal of Finance*, *45*(4), 1019-1044.

Marsh, P. (1982). The choice between equity and debt. *Journal of Finance*, *37*(1), 121-145.

Modigliani, F., & Miller, M. (1963). Corporate income taxes and the cost of capital. *American Economic Review*, *53*(3), 433-443.

Myers, S., & Majluf, H. (1984). Corporate financing and investment decisions when firms have information that investors do not have. *Journal of Financial Economics*, *13*(2), 187-221.

Myers, S. C. (1984). The capital structure puzzle. *The Journal of Finance*, *39*(3), 575-592.

[1] We report the correlation between beta and debt ratio for the three years to be .07, .22, and .26 for 2001, 2002, and 2003, respectively. Although positive for all three years (as it should be), the correlation was not as strong as expected (2001 in particular). It is unclear whether this is normal or indicative of a problem in the data for beta which would bias our results. Regardless, our results should not be taken as conclusive. A much more comprehensive study is needed before an unequivocal conclusion can be made.

Joseph Cheng, State University of New York at Binghamton is Associate Professor at Ithaca College teaching Security Analysis & Portfolio Management, Financial Analysis and Forecasting, Finance and Finance Markets & Institutions. His other research interests are Performance of Socially Responsible Funds, Closed End Funds and Optimal Debt.

Alka Bramhandkar, State University of New York at Binghamton is Associate Professor at Ithaca College teaching Case Studies in Financial Management, Speculative Markets, Business Finance, and Financial Markets & Institutions. Professor Bramhandkar's research interests are Performance of Socially Conscious Mutual Funds & Companies in the Dow Jones Sustainability Index and Optimal Debt.

APPENDIX
Second Order Condition

This appendix contains the steps showing that the second derivative of (7) with respect to debt is negative, which indicates that the first order condition is consistent with value maximization. The superscript prime (′) indicates that the variable is in terms of a derivative with respect to debt. For example, B′ would be the first derivative of beta with respect to debt and B″ would be the second derivative of beta with respect to debt.

Notations:

M = market risk premium = expected market return – risk free rate
$R_e' = M\,B'$ (derived by taking the derivative of the CAPM)
$R_e' = M\,B''$
$k = R_d/R_e$
$R_d' = k\,R_e' = k\,M\,B'$
$R_d'' = k\,R_e'' = k\,M\,B'$
$n = 1 - t$
$B' = B_u\,n\,A/E^2$
$B'' = 2\,B_u\,n\,A/E^3 = 2\,B'/E$
N = net income = $(EBIT - R_d\,D)\,n$ (1a)

(Note that N (capitalized) denotes net income whereas n (uncapitalized) denotes one minus the tax rate.)

The value of equity for the incumbent shareholders after the capital mix adjustment (7) can be expressed as:

$$PS_o = \frac{N + R_e\,d}{R_e} \tag{2a}$$

The first and second derivative of (1a) can be expressed as

$$PS_o' = (N' + R_e'\,d + R_e)\,R_e - R_e'\,(N + R_e\,d) \tag{3a}$$

$$PS_o'' = (N'' + R_e''\,d + R_e' + R_e')\,R_e + (N' + R_e'\,d + R_e)\,R_e' - R_e''\,(N + R_e\,d) - (N' + R_e'\,d + R_e)\,R_e' \tag{4a}$$

Since the denominator of the second derivative is Re^4, which is positive, we need only to write out the numerator of the second derivative in (4a). That is because a positive divisor does not affect the sign of the second derivative. The second and fourth terms in (4a) are identical so (4a) can further be simplified to :

$$= (N'' + R_e''\,d + R_e' + R_e')\,R_e - R_e''\,(N + R_e\,d)$$

$$= N''\,R_e + R_e''\,d\,R_e + 2\,R_e'\,R_e - R_e''\,N - R_e''\,R_e\,d \tag{5a}$$

The second and the last term in (5a) cancel out each other so that (5a) becomes:

$$= N''\,R_e + 2\,R_e'\,R_e - R_e''\,N \tag{6a}$$

N'' in (6a) represents the second derivative of N with respect to debt. To obtain its value, we begin with N (1a) and take its first and second derivatives:

$$N = (EBIT - R_d\,D)\,n$$

$$N' = -R_d'\,D\,n - R_d\,n$$

$$N'' = -R_d''\,D\,n - R_d'n - R_d'n$$

$$= -R_d''\,D\,n - 2\,R_d'n \tag{7a}$$

Substituting (7a) into (6a), we get

$$P\,S_o'' = (-R_d''\,D\,n - 2\,R_d'n)\,R_e + 2\,R_e'\,R_e - R_e''\,N \tag{8a}$$

Substituting terms from the notation section at the beginning of the appendix for R_d'', R_d', R_e' and R_e'' in (8a), we get

$$= (-k\,M\,B''\,D\,n - 2\,k\,M\,B'\,n)\,R_e + 2\,M\,B'\,R_e - M\,B''\,N$$

$$= 2\,M\,B'\,R_e\,(1 - k\,n) + M\,B''\,(-k\,D\,R_e\,n - N) \tag{9a}$$

Substituting B'' in (9a) with $2\,B'/E$, we get

$$= 2\,M\,B'\,R_e\,(1 - k\,n) + M\;2\,B'/E\,(-k\,D\,R_e\,n - N) \tag{10a}$$

Dividing both sides of (10a) by 2 M B′ (which is positive and thus does not affect the sign), we can express (10a) in a simple form without this positive divisor as:

$$= R_e (1- k n) - N/E - k R_e n D/E \tag{11a}$$

since $N/E = R_e$, we can write

$$= R_e (1- k n) - k R_e n D/E - R_e \tag{12a}$$

Dividing both sides of (12a) by R_e, we can express (12a) without the positive divisor as

$$= 1- k n - k n D/E -1$$

$$= - k n - k n D/E$$

Since k, n , and D/E are all positive, (13a) is clearly negative.
And because the second derivative is negative, the first derivative is indeed a maximum.

A numerical example showing that the solution is a maximum has been added to the body of the paper.

Table 2
Actual vs. Optimal Debt to Equity Ratios

	2001			2002			2003		
Company Symbol	Actual D/E	Optimum D/E	DIF	Actual D/E	Optimum D/E	DIF	Actual D/E	Optimum D/E	DIF
ED	0.64	1.01	(0.37)	0.65	0.81	(0.16)	0.67	0.74	(0.07)
CEG	0.62	0.87	(0.25)	1.01	1.23	(0.22)	0.77	1.02	(0.25)
D	0.84	1.13	(0.29)	0.80	0.89	(0.09)	0.76	1.04	(0.28)
DUK	0.45	0.79	(0.34)	1.21	1.14	0.07	1.09	1.29	(0.20)
EAS	1.27	1.42	(0.15)	1.16	1.13	0.03	1.22	1.27	(0.05)
EXC	0.84	1.08	(0.24)	0.77	0.99	(0.22)	0.62	0.79	(0.17)
FPL	0.49	0.97	(0.48)	0.53	0.57	(0.04)	0.72	0.75	(0.03)
AVA	2.02	1.27	0.75	1.81	1.54	0.27	1.19	1.24	(0.05)
BKH	0.46	0.69	(0.23)	0.87	0.79	0.08	0.91	0.90	0.01
EIX	2.58	0.60	1.98	2.99	1.85	1.14	1.65	0.90	0.75
EE	0.85	0.99	(0.14)	1.12	0.92	0.20	0.95	0.87	0.08
SRE	0.71	0.75	(0.04)	0.88	0.62	0.26	0.56	0.53	0.03
LNT	1.02	0.62	0.40	1.75	1.00	0.75	0.77	0.60	0.17
AEE	0.49	1.03	(0.54)	0.54	1.08	(0.54)	0.54	1.01	(0.47)
AEP	0.72	0.95	(0.23)	0.97	0.89	0.08	1.00	1.30	(0.30)
CIN	0.68	0.90	(0.22)	0.72	0.83	(0.11)	0.60	0.63	(0.03)
CNL	0.63	0.90	(0.27)	1.35	1.11	0.24	1.07	1.02	0.05
DPL	0.80	0.94	(0.14)	1.26	1.04	0.22	0.74	0.95	(0.21)
DTE	1.16	1.29	(0.13)	0.98	1.07	(0.09)	1.13	1.12	0.01
EDE	0.88	0.89	(0.01)	1.00	0.92	0.08	0.82	0.87	(0.05)
ETR	0.84	1.40	(0.56)	0.70	0.61	0.09	0.56	0.93	(0.37)
WEC	1.31	1.51	(0.20)	1.11	1.39	(0.28)	0.91	1.18	(0.27)
WPS	0.65	0.32	0.33	0.67	0.64	0.03	0.52	0.69	(0.17)
Average	0.91	0.97	(0.06)	1.08	1.00	0.08	0.86	0.94	(0.08)

INCOME INEQUALITY AND SECTOR SHIFT IN PENNSYLVANIA COUNTIES

Martina Vidovic, Bloomsburg University of Pennsylvania

ABSTRACT

This paper examines the impact of sector shift toward the reliance on the service sector for employment on income inequality in the state of Pennsylvania after accounting for other relevant factors hypothesized to influence income inequality. The results indicate that Pennsylvania counties associated with a greater percent age of employment in the service sector have significantly higher income inequality while counties with a higher percent age of population employed in manufacturing have lower income inequality. The results are robust to several different measures of income inequality that include the Gini coefficient, the Thiel and the Atkinson indices, and the coefficient of variation.

INTRODUCTION

Over the last few decades, social scientists have become increasingly interested in income inequality in the United States. The economic expansion of the 1980s and 1990s has not been shared equally by all segments of the population. While the average household income has been rising since late 1970s, the top 20% of households has been receiving an increasingly disproportionate amount of the aggregate household income. This has exacerbated the inequality in the distribution of income over time. The Census Bureau shows that, in 1980, the lowest 20% of households received 4.3% of the aggregate household income, while the top 20% received 43.7% of the total household income. However, by 1990, the share of aggregate income received by the bottom 20% fell to 3.9% and even further to 3.4% in 2004, while the share of aggregate income received by the top 20% of households increased to 46.6% by 1990 and to 50.1% by 2004.

Simon Kuznets (1955) proposed a theory that relates income inequality to economic development over time within a nation. He argued that there is a "bell-shaped" relationship (known as the inverted U-hypothesis) between income inequality and per capita income as a measure of economic development. According to the Kuznets' hypothesis, income inequality increases in the early stages of economic development (income growth) but declines in the later stages after a certain income threshold has been reached. In the early stages of development, inequality increases as increasing population growth hurts the poor and most of the wealth is held in the hands of a few entrepreneurial households. In the later stages of growth, the position of low income households improves as social and economic institutions emerge. The growth in income at the top of the income distribution slows down and ultimately income inequality declines. A similar view of the Kuznets' hypothesis (Cloutier, 1997) is that the rise in the manufacturing sector increases average incomes and reduces inequality through a convergence of skills. Subsequently, the decline in manufacturing and the growth of the service sector lead to a greater division among workers in terms of skills and, hence, an increase in inequality in the distribution of income. The change in the industrial composition in the U.S. seems to support this theory. According to census figures, the shares of employment in the manufacturing sector have declined from 21.7% in 1980 to 14.7% in 2000, while the shares of employment in the service sector have increased from 29.3% to 36.8% over the same period, and, at the same time, the inequality in the distribution of income has increased.

This paper examines the impact of the shift in employment from manufacturing to services on income inequality in the state of Pennsylvania after accounting for other factors hypothesized to influence income inequality. Manufacturing still accounts for a larger share of employment in Pennsylvania than in the nation (11.7 and 9.7%, respectively). Over time however, the state industrial structure has followed the national trend of a greater reliance on the service sector for employment and a lesser reliance on manufacturing. In 1990, slightly over 23% of the labor force in Pennsylvania was working in manufacturing and about 65% in services. In 2000, manufacturing employment was less than 20% while service sector employment increased to 71%. The regression results indicate that Pennsylvania counties with a greater percent age of employment in the service sector have significantly higher income inequality than counties with a higher percent age of employment in manufacturing, even after accounting for other relevant characteristics of a county: the level of median family income, population density, the percent age of farm population, educational attainment, the percent age of unemployed, the percent age of elderly population, the percent age of females in labor force, and the percent age of female-headed households.

LITERATURE REVIEW

The effect of economic restructuring associated with the decline of high wage manufacturing jobs and the expansion of low wage service jobs on the increase in income inequality observed over the past few decades has attracted substantial attention in research. However, the results thus far are inconclusive regarding the importance of economic restructuring on income inequality. Disagreeing conclusions can be partly attributed to different methodologies used, time periods covered, and geographic units used in analyses. Some researchers examined the inequality in earnings (Nelson & Lorence, 1988; Lorence, 1991; Morris & Western, 1999), while others studied the inequality in family or household income (Nielsen & Alderson, 1997; Partridge, Partridge & Rickman, 1998; Chevan & Stokes, 2000). Some used states as a level of analysis (Partridge et al., 1998; Morrill, 1998); others used counties or metropolitan areas (Lorence, 1991; Nielsen & Alderson, 1997; Chevan & Stokes, 2000; McLaughlin, 2002). More recently, the focus of research has been the change in income inequality over time (Chevan & Stokes, 2000; McLaughlin, 2002) rather than the level of inequality at a particular point in time.

Earlier studies (Bloomquist & Summers, 1982; Jacobs, 1985) found that the growth in the concentrated sector (corresponding to manufacturing sector) is negatively related to income inequality, while the growth in the competitive sector (corresponding to services) is positively related to income inequality. Recent studies that more directly examined the relationship between income inequality and the sector shift also found a positive relationship between a decline in manufacturing and/or increase in service employment and an increase in income inequality.

Nelson and Lorence (1988) found that the growth in service sector employment significantly increased metropolitan earnings' inequality between 1970 and 1980 after accounting for the effects of population characteristics. Lorence (1991), who primarily examined gender inequality, found that, although the growth in service sector employment reduced gender inequality, it increased the overall inequality by lowering the wages of males. The strongest effect was due to the growth in personal and social service employment.

Nielsen and Alderson (1997) found that manufacturing employment had a strong negative effect on income inequality across the U.S. counties in 1970, 1980 and 1990 after accounting for other aspects of economic development such as the level of income and socio-economic characteristics of a county. Morrill (2000) examined geographic variation in family income inequality across states from 1970 to 1990 and found that income inequality rose in states with high levels of service employment.

Partridge et al. (1998) examined the changes in the family income inequality across states between 1970 and 1990 and found that the growth in the employment in goods producing sectors which include construction, mining and manufacturing may have a potential for reducing income inequality. The relationship was negative but statistically significant in only one regression.

A recent study by Chevan and Stokes (2000) examined the change in family income inequality between 1970 and 1980 and between 1980 and 1990 and found that the effect of change in manufacturing employment was a stronger predictor of change in income inequality between 1970 and 1980 than between 1980 and 1990. Furthermore, the growth in employment in trade and personal services had the strongest effect on the increase in income inequality among all service sector employment particularly during the period 1980 to 1990. McLaughlin (2002) found that, although economic restructuring was an important determinant of an overall change in income inequality between 1980 and 1990, it was more important for explaining the change in income inequality among non-metro than metro counties.

In addition to industrial restructuring, the same studies also identified other factors that potentially contributed to the increase in income inequality which can be broadly classified as changes in labor supply and changes in demographic and household composition. The most important include changes in age and family structure, racial composition, and returns to education.

This study adds to the existing debate on the effect of economic restructuring on income inequality by examining the effect of manufacturing and service sector employment on family income inequality in Pennsylvania while using the most recent census data and accounting for other relevant factors identified in the existing literature.

MEASURES OF INCOME INEQUALITY

To examine income distribution in the 67 counties in Pennsylvania, this study uses data on gross family income from the 2000 Census Summary Tape File STF3C. Because Census reports income in categories, the income distributions are based on the number of families that fall in each income category. It is assumed that each family earns the midpoint of the income interval, except for the open-ended "$200,000 or more" category where an adjustment based on fitting a Pareto distribution is used to determine the average income earned by this category (Klein, 1962, pp. 150-154; Parker & Fenwick, 1983). The four measures of income inequality employed here include the coefficient of variation (CV), the Gini coefficient, the Thiel entropy index, and the Atkinson deprivation index. The last three measures are briefly described below.

The Gini coefficient derives from the Lorenz curve which plots the relationship between the cumulative income shares and the cumulative population shares ranked by income from the lowest to the highest. The Gini coefficient is then calculated as a ratio of the area between the actual income distribution (Lorenz curve) and the diagonal line representing perfect equality to the total area below the diagonal. The value of the coefficient ranges from a maximum of 1 representing total inequality to 0 representing perfect equality. When the income category data are used, the Gini coefficient, G, is computed as

$$G = 1 - \sum_{i=1}^{n} f_i(p_i + p_{i-1}) \qquad (1)$$

where f_i is the proportion of families in income category i and p_i is the proportion of total income received by families in income category i and all lower income categories.

Thiel proposed a measure of inequality based on information entropy. Applied to income distribution, the Thiel index of income inequality, T, is calculated as

$$T = \sum_{i}^{n}\left[\left(\frac{p_i}{p}\right) * \left(\frac{y_i}{\mu}\right) * \ln\left(\frac{y_i}{\mu}\right)\right] \qquad (2)$$

where p_i is the population of income group i, p is the total population, y_i is the average income in group i and μ is the average income in the county. A larger value of this index indicates greater inequality.

The Atkinson index is one of a few inequality measures that incorporate the social welfare function. The Atkinson index, A, is given by

$$A = 1 - {}^{y_\varepsilon}\!/_{\mu} \qquad (3)$$

where, μ is the mean income in the county and y_ε is the equity-sensitive average income defined as the level of per capita income that, if enjoyed by everyone, would make the total welfare exactly equal to the total welfare generated by the actual income distribution (Atkinson, 1970). The equity-sensitive average income (y_ε) is given by the expression

$$y_\varepsilon = \left[\sum_{i}^{n} p_i y_i^{1-\varepsilon}\right]^{\frac{1}{1-\varepsilon}} \qquad (4)$$

where p_i is the proportion of families in the ith income category, y_i is the average income in the ith income group, ε is a parameter that reflects society's aversion toward inequality, and n is the total number of income groups. The inequality aversion parameter ε can range from zero to infinity. As ε increases from zero, more weight is attached to income transfers at the lower end of the distribution and less to transfers at the top. The larger the parameter, the greater is the society's aversion toward inequality. Choosing ε is a normative decision and investigators should use several values of inequality aversion. Two values of ε used in this study (also used by Atkinson) are 0.5 giving emphasis to inequality at the top and 2 giving emphasis to inequality at the bottom of the distribution. For any income distribution, the value of A is between 0 and 1, where lower values indicate more equal income distribution with y_ε closer to μ.

Most of the empirical research on income inequality focuses on the Gini coefficient, despite some of its weaknesses. For example, economies can have similar average incomes but different income distributions and shapes of Lorenz curves and still have the same Gini coefficients. The same limitation exists when Lorenz curves of economies under comparison intersect (Allison, 1978). The Gini coefficient is more sensitive to the changes in income of the middle class than to that of the extremes (Allison, 1978). Braun (1988) concludes that this explains why Gini ratios show stability of income inequality in the U. S. over long periods of time. In addition, the Gini coefficient is influenced by the method used to compute it. Often, the Gini ratio is calculated using the Census income categories with an adjustment based on the Pareto curve for the open-ended category. The Pareto curve adjustment is not always done, and the number of income categories reported differs from one Census year to another which makes comparisons over time inconsistent.

The Gini coefficient will be lower as the number of income categories is reduced (Sale, 1974). As a result of these criticisms, other measures (for example the Atkinson and the Theil indices) are used in conjunction with the Gini coefficient.

Table 1 (page 37-38) compares the scores and the rankings of the four measures of income inequality among the 67 counties in Pennsylvania. Atkinson (1970) in his study of distribution of income found the rankings between the Gini coefficient and his measure to be more similar when the aversion parameter ε is less than 1 than when ε is 2. For the county data in this study, 21 out of 67 rankings are exactly the same comparing the Gini and the Atkinson index with $\varepsilon = 0.5$, and only 10 are the same comparing the Gini and the Atkinson index with $\varepsilon = 2$. If we consider rankings that only differ for up to two places, 46 out of 67 rankings compare between the Gini and the Atkinson index with $\varepsilon = 0.5$, and 20 out of 67 rankings compare between the Gini and the Atkinson with $\varepsilon = 2$. The correlations among the four inequality measures shown in Table 2 (page 38) are high and statistically significant at the 1% level. The strongest correlation of 0.9916 is between the Gini and the Atkinson index with $\varepsilon = 0.5$, while the lowest correlation of 0.6152 is between the coefficient of variation and the Atkinson index with $\varepsilon = 2$.

SOURCES OF INCOME INEQUALITY

The main hypothesis in this study is that income inequality in Pennsylvania is directly related to the sector shift from manufacturing to services. This hypothesis is tested using variables that capture the percent age of population employed in manufacturing (MANUF) and the percent age of population employed in the service (SERVICE) sector. Since wages in the service sector tend to be, on average, lower than in the manufacturing sector, and the workers tend to be more diverse in terms of skills in the service sector, it is expected that income inequality will be higher in the service sector and lower in the manufacturing sector.

The Kuznets' hypothesis suggests a negative relationship between income inequality and the level of economic development. The median family income (MEDINC) captures the level of economic development in a county (since it is closest to the measure of income per capita), and it is hypothesized that there will be an inverse relationship between median family income and income inequality. Furthermore, Kuznets suggested that inequality will be greater in densely populated urban areas where social conditions are more diverse than in rural areas. This is captured by including a variable that measures the population density (POPDEN) of a county and the relationship between population density and income inequality is hypothesized to be positive.

According to Kuznets (1955), a natural increase in population (measured as the difference between birth rate and death rate) would exacerbate income inequality. The population growth tends to increase the supply of unskilled labor and gives more weight in the income distribution to a larger percent age of young population at the bottom of the earning scale. Today, most developed economies (including the U.S. counties) are at the stage where large declines in birth rate exceed the continued decline in death rate leading to an increase in the average age but to a smaller growth of population than in the past. Therefore, the effect of the rate of natural population increase (POPINCR) on income inequality is expected to be positive, although very small.

An additional explanation for the inverted-U pattern that Kuznets (1955) argued is reflected in the differences in income inequality in the agricultural and the non-agricultural sectors of a developing economy. Kuznets argued that inequality is lower in the agricultural sector represented by relatively equally sized production units than in the non-agricultural sector. As the size of the more equal agricultural sector declines over time, the decreasing contribution of this sector will lead to an increase in income inequality. The size of the agricultural sector is captured by the percent age of farm population (FARMPOP) in a county, and it is hypothesized that the size of the agricultural sector is positively related to income inequality.

The level of education tends to be highly correlated with the level of development (measured by the median family income); however, some researchers argue (Jacobs, 1985) that the dissemination of education in a society is closely related to the distribution of income. Greater dispersion of educational attainment in a society leads to greater dispersion of opportunities and, hence, greater dispersion of income and higher income inequality. Following Nielsen and Alderson (1997), the dispersion of education is captured with Thiel's entropy formula:

$$EDUC = \sum_{i=1}^{3} p_i \ln\left(\frac{1}{p_i}\right) \qquad (5)$$

where p_1, p_2 and p_3 are the proportions of adult population (ages 25 and older) without a high school degree, with a high school degree only, and with a four year college degree only, respectively. This variable takes on the maximum value when the adult population is equally distributed among the three categories of educational attainment and the minimum value when the entire population is represented by a single category. The hypothesized effect of this variable on income inequality is positive.

Additional factors that may have led to the rise in inequality in family income and earnings that started in the 1970s include the changing role of women, the change in the socio-economic status of the elderly and the competition from low wage countries (Nielsen & Alderson, 1997). The participation by women (FEMALELF) in the labor force has been increasing over time along with the percent age of female-headed households (FEMALEHEAD). Females often earn below the average level of income; they tend to be paid less than men and more often work part-time. This, in addition to an increase in the percent age of female-headed households, tends to increase the proportion of low income earners in the income distribution. Therefore, a positive relationship between the percent age of females in labor force and the percent age of female headed households and income inequality is expected to be found.

On the other hand, since the early 1970s, elderly families have moved upward in the income distribution from the bottom to the middle. Their increases in income, coupled with improvements in health care, have led to an increase in the number of elderly families over time. For example, in Pennsylvania, the percent age of population older than 65 has increased by 1% between the 1990 and 2000 Census years. Therefore, it is hypothesized that there will be a negative relationship between income inequality and the percent age of population older than 65 (AGE65).

In addition, Wood (1994) argued that competition from countries with relatively cheap, low-skill labor has led to a reduced demand for unskilled labor in industrialized countries and to a greater difference in wages between skilled and unskilled workers. As the gap between relative wages of skilled and unskilled workers widens, the income inequality tends to widen as well. Lower demand for unskilled labor increases the unemployment rate, and the expected relationship between income inequality and the percent age of population that is unemployed (UNEMPL) in a county is positive. The definitions of the variables used in this study and the sources of the data are summarized in Table 3 (page 38).

The summary statistics in Table 4 (page 39) reveal that the income inequality in Pennsylvania counties is similarly dispersed around the mean when measured by the Gini coefficient and the Atkinson index with an inequality aversion parameter of 0.5. A slightly higher dispersion exists when income inequality is measured by the Thiel index. The largest dispersion around the mean exists in the case of the Atkinson index with the aversion parameter of 2. On average, in Pennsylvania counties, 16.2% of the population is older than 65 years of age; 19.6% of the labor force is employed in manufacturing while 71% is employed in the service sector. Farm population accounts for only 1.5% of population. Almost one half of the female population is in the labor force and 14% of all family households are female-headed households.

Pearson correlation coefficients shown in Table 5 (page 39) reveal a positive and statistically significant (at the 1% level) association between the percent age of population employed in the service sector and all measures of income inequality, as well as a negative and statistically significant correlation between the percent age of population employed in manufacturing and income inequality. The correlations between income inequality and the percent age of unemployed population, the percent age of female headed households, population density, and educational dispersion are strong and positive, while the correlations between income inequality and the percent age of females in labor force, and income inequality and the population increase are negative.

DISCUSSION OF RESULTS

This paper uses the ordinary least squares (OLS) regression analysis to estimate models with the four measures of income inequality as dependent variables. The median family income and population density are transformed into natural logs. For each dependent variable, three models are estimated. Model 1 includes only the Kuznets' original variables, while Models 2 and 3 include all the other socio-economic variables. The percent ages of population employed in manufacturing and in services are not included in the same model since the two are highly negatively correlated. Instead, two separate models are estimated (Model 2 versus Model 3). The results can be compared between models for each dependent variable and between measures of the dependent variable for each model. Tables 6 through 10 (pages 40-42) show the results. The coefficients in Tables 7 through 10 are not standardized and cannot be easily interpreted as such. However, the size of the coefficients can be compared across the models in each table. The focus in this paper is on the signs of the coefficients.

For each dependent variable, the explanatory power of the model increases (measured by R square) once other socio-economic variables are included with those in Model 1. All base models (Model 1) find a statistically significant and negative relationship between the median family income and income inequality, and a positive and statistically significant relationship between income inequality and population density as well as the dispersion of education in a county. These results are as anticipated and are consistent with the findings of some previous studies. Nielson and Alderson (1997) found the identical signs on the same variables. Partridge et. al. (1998) also found that an increase in per capita income and average years of education has a potential to reduce income inequality at a state level. Similarly, McLaughlin (2002) found a positive and statistically significant relationship between income inequality and the level of education, median household income, and the size of the population. Once all the other socioeconomic variables are included, the coefficient on the median family income is no longer statistically significant (Model 1 versus Models 2 and 3) except in the case where the dependent variable is the Atkinson index with the aversion parameter of 2, the coefficient is positive and statistically significant.

Models 2 and 3 provide a statistically significant support for the main hypothesis that income inequality increases with the size of the service sector but decreases with the size of the manufacturing sector. The coefficient on the percent age of labor force employed in manufacturing is negative and statistically significant in the regressions where the dependent variable is the Gini coefficient or the Atkinson index, while the coefficient on the percent age of labor force employed in services is positive and statistically significant in all regressions. These results validate the findings of studies discussed earlier.

In particular, these results show that, at least in Pennsylvania, the decline in the size of the manufacturing sector still has a statistically significant effect on the increase in income inequality. Recall that Chevan and Stokes (2000) found that the size of the manufacturing industry had an effect on the change in income inequality between 1970 and 1980 and no effect between 1980 and 1990, while Partridge et. al (1998) found only a weak relationship between the Gini coefficient and goods producing employment. If we view the shift in employment from manufacturing to services as an inevitable succession of a developed economy, policy efforts oriented toward reducing the income inequality should focus on educating the labor force rather than trying to recapture the success of the manufacturing industry.

To check the robustness of this result, a separate model was constructed with the log of the ratio of service to manufacturing employment. Because this ratio is always greater than 1, the log will be positive, and the positive coefficient on this variable would indicate that, as the ratio of service to manufacturing employment increases, so too the inequality in the distribution of income. Indeed the coefficient on this variable is positive, but it is only statistically significant at the 10% level when the dependent variable is the Gini coefficient or the Atkinson index. The results of this regression are not reported in the paper, but they are available from the author.

Across models in all tables, except Table 7 where the dependent variable is the coefficient of variation, income inequality is positively and statistically significantly related to the percent age of population that is unemployed and the percent age of female-headed households. Also, the coefficient on the percent age of farm population is positive and statistically significant in most models. All these coefficients have the expected signs and are consistent with the findings by Nielsen and Alderson (1997), Partridge et al. (1998), Chevan and Stokes (2000) and McLaughlin (2002).

However, this study finds that income inequality is lower in counties with a higher proportion of females in labor force which is contrary to expectations. An equivalent finding was also reported by Nielsen and Alderson (1997) who analyzed income inequality across all counties in the United States for census years 1970, 1980 and 1990. Chevan and Stokes (2000) also found that an increase in wives' labor force participation led to a decrease in income inequality but only in the period 1980 to 1990. Women's sharp rise in labor force participation over the last few decades, coupled with fact that they traditionally have lower levels of experience and lower wages, seemed a prime candidate for explaining the increase in income inequality.

The result of this study, in contrast, leads to the conclusion that an increase in female labor force participation increases the absolute level of income of lower income households (both single and two-parent) without depressing the wages of male workers, thus leading to a reduction in income inequality. This indicates that public policy oriented toward reducing income inequality should focus on programs that would ease women's effort in participating in the labor force such as providing the adequate child care.

Another factor that led to a change in income inequality is the change in the size of elderly population. The coefficient on the percent age of population older than 65 years of age is negative and statistically significant only when the dependent variable is the Atkinson index and the aversion parameter is 2. Thus, the evidence that income inequality decreases with an increase in elderly population is rather weak. A possible explanation for this result is that families headed by an elderly person are still in the lower income group (the correlation coefficient between median household income and the percent age of population older than 65 years is -.5 and it is statistically significant at the 1% level. Chevan and Stokes (2000) found that the size of the elderly population tended to increase inequality in metropolitan areas during the 1970-1980 period but had a negative, albeit insignificant, effect during the 1980-1990 period. Similarly, Partridge et. Al. (1998) found a negative but statistically insignificant effect of the percent age of elderly population on state income inequality. This suggests that the public policy could bring about a reduction in income inequality by focusing on improving the position of the growing elderly population.

CONCLUSIONS

Income inequality in the United States and in Pennsylvania alone has been increasing in the past three decades. This paper examines the relationship between income inequality and sector employment in Pennsylvania counties using various measures of income inequality and controlling for other important socio-economic factors. The results indicate that an increase in income inequality is positively related to a shift in employment from the manufacturing sector to the service sector, even after controlling for the influence of all other factors in the model. The results are robust across different measures of income inequality. The highest similarity in results exists when the department variables are the Gini coefficient and the Atkinson index with a low level of aversion to inequality followed by the Thiel index.

The results of this study provide broad implications for the pubic policy where the efforts are oriented toward reducing the income inequality. Given the direction of economic development toward greater reliance on the service sector for employment and less on the manufacturing sector, the policies aiming at reducing income inequality should focus on improving the educational attainment of the labor force, improving the opportunity for women in participating in the labor force, and improving the conditions of the growing elderly population.

REFERENCES

Allison, P. D. (1978). Measures of inequality. *American Sociological Review*, *43*(6), 865-880.

Atkinson, A. (1970). On the measurement of inequality. *Journal of Economic Theory*, *2*(3), 244-263.

Braun, D. (1988). Multiple measurements of U.S. income inequality. *The Review of Economics and Statistics*, *70*(3), 398-405.

Bloomquist, L., & Summers, G. (1982). Organization of production and community income distribution. *American Sociological Review*, *47*(3), 325-338.

Chevan, A., & Stokes, R. (2000). Growth in family income inequality, 1970-1990:Industrial restructuring and demographic change. *Demography*, *37*(3), 365-380.

Cloutier, N. (1997). Metropolitan income inequality during the 1980s: The impact of urban development, industrial mix, and family structure. *Journal of Regional Science*, *37*(3), 459-478.

Jacobs, D. (1985). Unequal organizations or unequal attainments? an empirical comparison of sectoral and individualistic explanations for aggregate inequality. *American Sociological Review*, *50*(2), 166-180.

Klein, L. R. (1962). *An introduction to econometrics*. Englewood Cliffs, NJ: Prentice Hall.

Kuznets, S. (1955). Economic growth and income inequality. *The American Economic Review*, *65*(1), 1-28.

Lorence, J. (1991). Growth in service sector employment and MSA gender earnings inequality. *Social Forces*, *69*(3), 763-783.

McLaughlin, D. K. (2002). Changing income inequality in nonmetropolitan counties, 1980 to 1990. *Rural Sociology*, *67*(4), 512-533.

Morrill, R. (2000). Geographic variation in change in income inequality among US states, 1970-1990. *The Annals of Regional Science*, *34*(1), 109-130.

Morris, M., & Western, B. (1999). Inequality in earnings at the close of the twentieth century. *Annual Review of Sociology*, *25*, 623-657.

Nelson, J., & Lorence, J. (1988). Metropolitan earning inequality and service sector employment. *Social Forces*, *67*(2), 492-511.

Nielsen, F., & Alderson, A. S. (1997). The Kuznets curve and the great u-turn: Income inequality in U.S. counties, 1970-1990. *American Sociological Review*, *62*(1), 12-33.

Parker, R. N., & Fenwick, R. (1983). The Pareto curve and its utility for open-ended income distributions in survey research. *Social Forces*, *61*(3), 872-885.

Partridge, J. S., Partridge, M. D., & Rickman D. S. (1998). State patterns in family income inequality. *Contemporary Economic Policy*, *16*(3), 277-294.

Sale, T. S. (1974). Interstate analysis of the size distribution of family income 1950-1970. *Southern Economic Journal*, *40*(3), 434-441.

Wood, A. (1994). *North-South trade, employment and inequality: Changing fortunes in a skill-driven world.* Oxford, England: Clarendon Press.

Martina Vidovic, Binghamton University, is Assistant Professor of Economics at Bloomsburg University teaching Principles of Economics, Micro and Managerial Economics, Economics, Environmental Economics, Business and Economics Statistics and Business and Economics Math. Her other research interests are voluntary pollution prevention programs, distribution of pollution, income inequality and health.

Table 1
Family Income Inequality and Relative Rankings in Pennsylvania Counties, 2000

County	Gini	CV	ε=2.0 Atkinson	ε=0.5 Atkinson	Thiel
Elk	0.3137(1)	0.6252(1)	0.3828(1)	0.0874(1)	0.1736(1)
Adams	0.3211(2)	0.6308(2)	0.3956(2)	0.0904(2)	0.1793(2)
Cameron	0.3252(3)	0.6518(3)	0.4039(3)	0.0933(3)	0.1858(3)
Perry	0.3295(4)	0.6798(5)	0.4119(4)	0.0964(4)	0.1944(4)
York	0.3364(5)	0.6873(7)	0.4219(9)	0.0994(5)	0.2010(5)
Fulton	0.3392(6)	0.6937(8)	0.4243(10)	0.1012(9)	0.2038(7)
Carbon	0.3397(7)	0.6980(10)	0.4196(8)	0.1010(7)	0.2047(8)
Franklin	0.3401(8)	0.7063(16)	0.4170(7)	0.1011(8)	0.2062(9)
Pike	0.3416(9)	0.6767(4)	0.4254(11)	0.1007(6)	0.2016(6)
Warren	0.3420(10)	0.7096(17)	0.4138(5)	0.1019(10)	0.2081(10)
Lancaster	0.3438(11)	0.7042(15)	0.4365(17)	0.1034(11)	0.2096(12)
Juniata	0.3466(12)	0.7608(36)	0.4156(6)	0.1058(16)	0.2213(22)
Huntingdon	0.3473(13)	0.7178(21)	0.4355(16)	0.1057(15)	0.2139(15)
Lebanon	0.3482(14)	0.7232(23)	0.4328(13)	0.1056(14)	0.2156(17)
Beaver	0.3487(15)	0.7006(13)	0.4555(33)	0.1068(17)	0.2130(13)
Cumberland	0.3492(16)	0.7103(19)	0.4271(12)	0.1045(13)	0.2133(14)
Somerset	0.3494(17)	0.7303(25)	0.4344(14)	0.1072(8)	0.2183(20)
Bucks	0.3499(18)	0.6809(6)	0.4492(26)	0.1044(12)	0.2082(11)
Monroe	0.3523(19)	0.6959(9)	0.4590(36)	0.1078(9)	0.2144(16)
Northumberland	0.3530(20)	0.7519(30)	0.4422(20)	0.1098(22)	0.2251(26)
Northampton	0.3545(21)	0.7033(14)	0.4643(43)	0.1089(21)	0.2174(19)
Schuylkill	0.3555(22)	0.7627(37)	0.4407(9)	0.1110(24)	0.2291(27)
Berks	0.3556(23)	0.7139(20)	0.4697(49)	0.1103(23)	0.2208(21)
Forest	0.3556(24)	0.6987(11)	0.4443(22)	0.1089(20)	0.2168(18)
Mifflin	0.3559(25)	0.7648(40)	0.4470(24)	0.1117(26)	0.2298(29)
Wyoming	0.3591(26)	0.7096(18)	0.4591(38)	0.1111(25)	0.2217(23)
Bedford	0.3596(27)	0.8103(60)	0.4354(15)	0.1141(33)	0.2411(42)
Clearfield	0.3611(28)	0.7464(27)	0.4446(23)	0.1125(29)	0.2296(28)
Armstrong	0.3611(29)	0.7526(31)	0.4539(32)	0.1135(30)	0.2311(30)
Butler	0.3612(30)	0.7181(22)	0.4686(48)	0.1123(28)	0.2250(25)
Jefferson	0.3621(31)	0.7812(52)	0.4383(8)	0.1137(31)	0.2368(35)
Columbia	0.3628(32)	0.7761(48)	0.4441(2)	0.1141(34)	0.2367(34)
Chester	0.3636(33)	0.7002(12)	0.4757(55)	0.1118(27)	0.2226(24)
Tioga	0.3638(34)	0.7554(33)	0.4480(25)	0.1140(32)	0.2333(32)
McKean	0.3643(35)	0.7722(43)	0.4609(40)	0.1162(38)	0.2381(39)
Susquehanna	0.3663(36)	0.7635(38)	0.4538(30)	0.1158(36)	0.2372(38)
Potter	0.3667(37)	0.7691(42)	0.4527(28)	0.1163(39)	0.2388(40)
Lawrence	0.3674(38)	0.7439(26)	0.4751(54)	0.1171(41)	0.2357(33)
Venango	0.3676(39)	0.7787(51)	0.4721(50)	0.1186(46)	0.2424(44)
Crawford	0.3676(40)	0.7567(34)	0.4595(39)	0.1161(37)	0.2369(36)
Mercer	0.3685(41)	0.7741(46)	0.4679(47)	0.1180(43)	0.2421(43)
Centre	0.3687(42)	0.7518(29)	0.4675(46)	0.1165(40)	0.2371(37)
Union	0.3693(43)	0.8084(59)	0.4538(31)	0.1185(45)	0.2497(52)
Montgomery	0.3694(44)	0.7280(24)	0.4671(44)	0.1148(35)	0.2322(31)

Blair	0.3698(45)	0.7780(50)	0.4675(45)	0.1189(47)	0.2439(47)
Erie	0.3701(46)	0.7754(47)	0.4749(53)	0.1192(48)	0.2440(48)
Lycoming	0.3702(47)	0.7912(53)	0.4533(29)	0.1181(44)	0.2454(49)
Bradford	0.3708(48)	0.7737(44)	0.4727(51)	0.1193(49)	0.2438(46)
Westmoreland	0.3709(49)	0.7643(39)	0.4619(42)	0.1176(42)	0.2409(41)
Clarion	0.3722(50)	0.7764(49)	0.4801(58)	0.1208(52)	0.2458(50)
Clinton	0.3730(51)	0.8008(57)	0.4612(41)	0.1204(50)	0.2498(53)
Wayne	0.3742(52)	0.8169(62)	0.4574(35)	0.1213(55)	0.2544(59)
Luzerne	0.3746(53)	0.7739(45)	0.4762(56)	0.1209(53)	0.2465(51)
Lehigh	0.3754(54)	0.7498(28)	0.4930(61)	0.1207(51)	0.2427(45)
Montour	0.3778(55)	0.7998(56)	0.4503(27)	0.1209(54)	0.2534(58)
Indiana	0.3784(56)	0.7943(54)	0.4762(57)	0.1232(56)	0.2531(57)
Dauphin	0.3792(57)	0.7689(41)	0.5063(62)	0.1244(57)	0.2507(55)
Washington	0.3811(58)	0.7994(55)	0.4815(59)	0.1246(58)	0.2571(60)
Delaware	0.3820(59)	0.7591(35)	0.5141(64)	0.1248(59)	0.2505(54)
Greene	0.3837(60)	0.7554(32)	0.5128(63)	0.1277(62)	0.2523(56)
Snyder	0.3838(61)	0.9585(67)	0.4590(37)	0.1318(64)	0.2930(65)
Lackawanna	0.3847(62)	0.8158(61)	0.4835(60)	0.1270(61)	0.2632(62)
Cambria	0.3871(63)	0.9137(66)	0.4744(52)	0.1317(63)	0.2844(63)
Sullivan	0.3882(64)	0.8012(58)	0.4556(34)	0.1254(60)	0.2599(61)
Fayette	0.4068(65)	0.8787(65)	0.5237(65)	0.1425(66)	0.2951(66)
Allegheny	0.4076(66)	0.8517(64)	0.5374(66)	0.1418(65)	0.2921(64)
Philadelphia	0.4312(67)	0.8815(67)	0.5875(67)	0.1601(67)	0.3220(67)

Table 2
Correlations Among Measures of Income Inequality

	Gini	**CV**	**ε=2 Atkinson**	**ε=0.5 Atkinson**
CV	0.8406[a]			
Atkinson ε=2	0.9103[a]	0.6152[a]		
Atkinson ε=0.5	0.9916[a]	0.8710[a]	0.9078[a]	
Thiel	0.9702[a]	0.9425[a]	0.8288[a]	0.9843[a]

[a] indicates statistically significant at the 1% level

Table 3
Variable Definitions and Sources

Variable	**Definition and Source**
	US Census 2000
MEDINC	Median family income
POPDEN	Population per square mile
UNEMPL	% civilian labor force (16+) that is unemployed
MANUF	% population employed in manufacturing
SERVICE	% population employed in wholesale, retail trade, service industry
EDUC	Thiel's entropy index (groups include: no high school degree, high school degree only and bachelor's degree)
FARMPOP	% farm population
FEMALEHEAD	% female headed family households
FEMALELF	(Females 16 + in labor force / total females 16+)*100
AGE65	(Population 65+ / total population)*100
	City and County Book 2000
POPINCR	Birth rate – death rate (1997 values used for 2000)

Table 4
Summary Statistics of All Variables

Variable	Mean	Standard Deviation
GINI	0.3629	0.0198
CV	0.7545	0.0606
ATKINSON, ε = 2	0.4571	0.0331
ATKINSON, ε = 0.5	0.1145	0.0120
THIEL	0.2339	0.0268
MEDINC	44806.81	8426.99
FARMPOP	1.4746	1.2879
POPDEN	453.16	1415.00
POPINCR	0.3567	3.0387
EDUC	0.9021	0.0343
AGE65	16.1866	2.2653
MANUF	19.6448	6.8294
SERVICE	71.0537	7.4811
UNEMPL	3.3209	0.7739
FEMALELF	53.5702	4.0826
FEMALEHEAD	13.8179	3.8630

Table 5
Correlations of Income Inequality with Socioeconomic Variables

	Inequality Measures				
Socio-economic Variables	Gini	Thiel	CV	Atkinson ε = 0.5	Atkinson ε = 2
MEDINC	-0.1520	-0.2553 [b]	-0.3686	0.0307	-0.2092 [c]
FARMPOP	-0.1626	-0.0492	0.1322	-0.3581 [a]	-0.1399
MANUF	-0.5168 [a]	-0.4270 [a]	-0.2735 [b]	-0.5755 [a]	-0.4831 [a]
SERVICE	0.5207 [a]	0.4243 [a]	0.2480 [b]	0.6248 [a]	0.4888 [a]
UNEMPL	0.5730 [a]	0.5393 [a]	0.4291 [a]	0.5034 [a]	0.5803 [a]
FEMALELF	-0.3202 [a]	-0.3622 [a]	-0.4022 [a]	-0.1827	-0.3456 [a]
FEMALEHEAD	0.6046 [a]	0.5589 [a]	0.3685 [a]	0.7338 [a]	0.6418 [a]
POPDEN	0.4820 [a]	0.4352 [a]	0.2489 [b]	0.5923 [a]	0.5130 [a]
POPINCR	-0.2635 [b]	-0.2555 [b]	-0.2630 [b]	-0.0709	-0.2401 [c]
AGE65	0.1398	0.1728	0.2260 [c]	-0.0182	0.1418
EDUC	0.2614 [b]	0.1855	0.0410	0.4127 [a]	0.2348 [c]

[a] indicates statistically significant at the 1% level;
[b] indicates statistically significant at the 5% level;
[c] indicates statistically significant at the 10% level

Table 6
Regression Results where the Dependent Variable is the Gini Coefficient, 2000

Variable	Model 1	Model 2	Model 3
MEDINC (LN)	-0.0710 [a]	0.0389	0.0386
	(0.0165)	(0.0269)	(0.0265)
FARMPOP	0.0012	0.0063 [a]	0.0073 [a]
	(0.0017)	(0.0017)	(0.0017)
POPDEN (LN)	0.0095 [a]	-6.81E-05	-0.0009
	(0.0024)	(0.0033)	(0.0033)
POPINCR	-0.0021 [a]	-0.0014	-0.0013
	(0.0008)	(0.0009)	(0.0009)
EDUC	0.2121 [b]	0.1603 [b]	0.1557 [b]
	(0.0888)	(0.0699)	(0.0687)
AGE65	-	-0.0007	-0.0008
	-	(0.0012)	(0.001)
MANUF	-	-0.0007 [b]	-
	-	(0.0003)	-
SERVICE	-	-	0.0008 [a]
	-	-	(0.0003)
UNEMPL	-	0.0079 [a]	0.0073 [a]
	-	(0.0028)	(0.0027)
FEMALELF	-	-0.0022 [a]	-0.0023 [a]
	-	(0.0007)	(0.0007)
FEMALEHEAD	-	0.0026 [a]	0.0026 [a]
	-	(0.0009)	(0.0009)
INTERCEPT	0.8820 [a]	-0.1281	-0.1826
	(0.1614)	(0.2637)	(0.2596)
N	67	67	67
R^2	0.4996	0.7252	0.7343
Adjusted R^2	0.4586	0.6762	0.6868

Standard deviation in parentheses.
[a] indicates statistically significant at the 1% level;
[b] indicates statistically significant at the 5% level;
[c] indicates statistically significant at the 10% level

Table 7
Regression Results where the Dependent Variable is the Coefficient of Variation, 2000

Variable	Model 1	Model 2	Model 3
MEDINC (LN)	-0.2519 [a]	-0.1064	-0.1111
	(0.0526)	(0.1085)	(0.1061)
FARMPOP	0.0133 [b]	0.0200 [a]	0.0228 [a]
	(0.0055)	(0.0069)	(0.0069)
POPDEN (LN)	0.0277 [a]	0.0202	0.0146
	(0.0076)	(0.0132)	(0.0132)
POPINCR	-0.0045 [c]	-0.0013	-0.0006
	(0.0024)	(0.0035)	(0.0034)
EDUC	0.4668	0.4157	0.4099
	(0.2829)	(0.2817)	(0.2752)
AGE65	-	0.0015	0.0025
	-	(0.0047)	(0.0045)
MANUF	-	-0.0012	-
	-	(0.0012)	-
SERVICE	-	-	0.0024 [c]
	-	-	(0.0013)
UNEMPL	-	0.01722	0.0144
	-	(0.0113)	(0.0111)
FEMALELF	-	-0.0042	-0.0039
	-	(0.0029)	(0.0028)
FEMALEHEAD	-	0.0005	0.0001
	-	(0.0036)	(0.0035)
INTERCEPT	2.8702 [a]	1.5438	.4085
	(0.5142)	(1.0626)	(1.0400)
N	67	67	67
R^2	0.4553	0.5212	0.542
Adjusted R^2	0.407	0.4357	0.4603

Standard deviation in parentheses.
[a] indicates statistically significant at the 1% level;
[b] indicates statistically significant at the 5% level;
[c] indicates statistically significant at the 10% level

Table 8
Regression Results where the Dependent Variable is the Atkinson Index, $\varepsilon = 2$, 2000

Variable	Model 1	Model 2	Model 3
MEDINC (LN)	-0.1159 [a]	0.0932 [b]	0.0937 [b]
	(0.0262)	(0.0381)	(0.0383)
FARMPOP	-0.003	0.0062 [b]	0.0073 [a]
	(0.0028)	(0.0024)	(0.0025)
POPDEN (LN)	0.0171 [a]	-0.0020	-0.0023
	(0.0038)	(0.0046)	(0.0048)
POPINCR	-0.0013	-0.001	-0.0012
	(0.0012)	(0.0013)	(0.0012)
EDUC	0.3139 [b]	0.2169 [b]	0.2098 [b]
	(0.1409)	(0.0989)	(0.0995)
AGE65	-	-0.0030 [c]	-0.0034 [b]
	-	(0.0016)	(0.0016)
MANUF	-	-0.0009 [b]	-
	-	(0.0004)	-
SERVICE	-	-	0.0009 [b]
	-	-	(0.0005)
UNEMPL	-	0.0085 [b]	0.0085 [b]
	-	(0.0039)	(0.0039)
FEMALELF	-	-0.0045 [a]	-0.0048 [a]
	-	(0.0010)	(0.0010)
FEMALEHEAD	-	0.0059 [a]	0.0058 [a]
	-	(0.0013)	(0.0013)
INTERCEPT	1.3329 [a]	-0.5341	-0.6019
	(0.2562)	(0.3733)	(0.3758)
N	67	67	67
R^2	0.5468	0.8019	0.7997
Adjusted R^2	0.5097	0.7665	0.7639

Standard deviation in parentheses.
[a] indicates statistically significant at the 1% level;
[b] indicates statistically significant at the 5% level;
[c] indicates statistically significant at the 10% level

Table 9
Regression Results where the Dependent Variable is the Atkinson Index, $\varepsilon = 0.5$, 2000

Variable	Model 1	Model 2	Model 3
MEDINC (LN)	-0.0509 [a]	0.0147	0.0144
	(0.0096)	(0.0158)	(0.0155)
FARMPOP	0.00007	0.0037 [a]	0.0042 [a]
	(0.0010)	(0.0010)	(0.0010)
POPDEN (LN)	0.0063 [a]	0.0004	-6.2E-05
	(0.0014)	(0.0019)	(0.0019)
POPINCR	-0.0009 [b]	-0.0006	-0.0006
	(0.0004)	(0.0005)	(0.0005)
EDUC	0.1192 [b]	0.0880 [b]	0.0866 [b]
	(0.0515)	(0.0410)	(0.0403)
AGE65	-	-0.0005	-0.0005
	-	(0.0007)	(0.0006)
MANUF	-	-0.0003 [b]	-
	-	(0.0002)	-
SERVICE	-	-	0.0005 [b]
	-	-	(0.0002)
UNEMPL	-	0.0043 [b]	0.0040 [b]
	-	(0.0016)	(0.0016)
FEMALELF	-	-0.0013 [a]	-0.0014 [a]
	-	(0.0004)	(0.0004)
FEMALEHEAD	-	0.0016 [a]	0.0016 [a]
	-	(0.0005)	(0.0005)
INTERCEPT	0.5196 [a]	-0.0807	-0.1094
	(0.0937)	(0.1547)	(0.1523)
N	67	67	67
R^2	0.5391	0.743	0.7499
Adjusted R^2	0.5014	0.6951	0.7052

Standard deviation in parentheses.
[a] indicates statistically significant at the 1% level;
[b] indicates statistically significant at the 5% level;
[c] indicates statistically significant at the 10% level

Table 10
Regression Results where the Dependent Variable is the Thiel Index, 2000

Variable	Model 1	Model 2	Model 2
MEDINC (LN)	-0.1122[a]	0.0102	0.0091
	0.0220	0.0403	0.0395
FARMPOP	0.0033	0.0090[a]	0.0102[a]
	0.0023	0.0026	0.0026
POPDEN (LN)	0.0137[a]	0.0039	0.0021
	0.0032	0.0049	0.0049
POPINCR	-0.0023[b]	-0.0013	-0.0011
	0.0010	0.0013	0.0013
EDUC	0.2603[b]	0.2063[c]	0.2019[c]
	0.1185	0.1047	0.1024
AGE65	-	-0.0005	-0.0003
	-	0.0017	0.0017
MANUF	-	-0.0007	-
	-	0.0004	-
SERVICE	-	-	0.0011[b]
	-	-	0.0005
UNEMPL	-	0.0094[b]	0.0086[c]
	-	0.0042	0.0041
FEMALELF	-	-0.0026[b]	-0.0026[b]
	-	0.0011	0.0011
FEMALEHEAD	-	0.0025[c]	0.0025[c]
	-	0.0013	0.0013
INTERCEPT	1.1255[a]	0.0015	-0.0639
	0.2154	0.3951	0.3868
N	67	67	67
R^2	0.5118	0.6617	0.6764
Adjusted R^2	0.4718	0.6013	0.6186

Standard deviation in parentheses.
[a] indicates statistically significant at the1% level;
[b] indicates statistically significant at the 5% level;
[c] indicates statistically significant at the10% level

WHAT DOES THE ACCOUNTING PROFESSION REALLY WANT?

Thomas W. Oliver, Clarion University of Pennsylvania
Charles J. Pineno, Shenandoah University

ABSTRACT

In a challenging global environment, employers are always looking for accounting professionals who have the requisite technical, analytical, communication and social skills. The adequacy of accounting curricula in meeting students' needs has been under scrutiny, with some in the profession even suggesting that accounting curricula lack relevance and that students' general and technical skills are not being developed adequately. The focus of this research is serious consideration of the employers' perspective on the development of accounting curricula. In this study we report the results of a recent survey of accounting alumni in regard to skills and preparation necessary for the profession, as well as data gathered over the past 10 years from supervisors' evaluations of accounting interns. We present response comparisons by gender, professional certification, job title and degree earned from the alumni survey. Skills and preparation sought by "Big Four," medium and small public accounting firms, corporations and governmental agencies will be a focus of our review of intern supervisors' evaluations. For comparison, evaluations are reviewed based upon the type of internship. Trends from other surveys provide another basis for comparison.

INTRODUCTION

In a challenging global environment, employers are always looking for accounting professionals who have the requisite technical, analytical, communication and social skills necessary to be successful. The need for young professionals who possess strong general and technical skills prompted the formation of the Accounting Education Change Commission (AECC) and the Bedford Committee of the American Accounting Association (AAA) (1986). Later position papers by the Institute of Management Accountants (IMA) (1994), and the American Institute of Certified Public Accountants (AICPA) (2002) also seemed to indicate that accounting curricula lacked relevance and that students' skills were not being developed adequately. While these charges are well formulated, we believe that an employer's perspective is critically important. The focus of this research is serious consideration of the employer's perspective on skills necessary for success and the development of accounting curricula.

RESEARCH METHOLOGY

In this study we report the results of a recent survey of accounting alumni about the skills and preparation necessary for the profession. Participants were selected as a random sample of 1,000 graduates from the Clarion University accounting program. Each was mailed a four page questionnaire during the fall of 2001. The response rate was 12.2% among all participants. Of those responding, 47% were male and 53% female. With regard to employment, 19% were currently employed in public accounting, 62% in industry accounting and 13% in other accounting jobs. A total of 32% were CPAs. Of the 19% employed in public accounting, 21% were in "Big Five" firms. In Table 1 (page 50), we summarize this data.

In addition, we also analyzed data gathered over the past 10 years from supervisors' evaluation of accounting interns. Skills and preparation sought by "Big Four," medium and small public accounting firms, corporations and governmental agencies were determined by a review of intern supervisors' evaluations. For comparison, evaluations were reviewed based upon the type of internship.

In the alumni survey, we considered broad categories of desired skills and knowledge, including communication, technology, personal skills, as well as technical accounting knowledge. We specifically asked about the importance of nine skill/knowledge criteria in pursuit of an accounting career. These were oral communication, written communication, interpersonal skills, computer and technology skills, decision making ability, ability to work independently, ability to work in a group, technical knowledge, and analytical thinking skills. All variables were measured using a 5-point Likert scale. We analyzed responses from the alumni survey by gender, certification, employer industry and, for public accounting, type of firm.

BACKGROUND STUDIES

Over the years the professional organizations have examined the needs of the accounting profession to help ensure that an accounting education prepares students for professional success. Groups such as the American Accounting Association (AAA), the American Institute of Certified Public Accountants (AICPA) and the Institute of Management Accountants (IMA) have done various studies and issued reports. Among these was a white paper of the then-current "Big Eight" accounting firms, **Perspectives on Education: Capabilities for Success in the Accounting Profession** (1989) which helped initiate the movement to change accounting education. The Accounting Education Change Commission (AECC) was then created by the AAA in 1989 to develop a "change of focus in the philosophy of accounting education." Many studies began to focus on skills and attributes of accounting graduates. The Institute of Management Accountants (IMA) study **Accountants: What Corporate America Wants in Entry-Level** (1994) broadened the view from the "big firm/public accounting" perspective to include the needs of industry as well.

Foster, Bolt-Lee and Colson (2002) summarize the CPA perspective, which derives from the AICPA's **Core Competency Framework for Entry into the Accounting Profession**: "The Framework was prepared for educators to develop a curriculum of broad-based accounting education, with a focus on the needs of the future professional accountant. When completed, the Framework will have three primary components. The first component defines a set of competencies that all students entering the field of accounting should possess. The task force made the decision to focus on competencies, or skills, rather than subject content or a body of knowledge because of the rapidly changing nature of the accounting profession. Students who possess a core set of competencies will be able to adapt to change and will be more valuable in the workplace, in turn opening more career opportunities to them."

In July 1999, the AICPA Board endorsed the competency framework with categories of competencies in three areas: functional, personal and broad business perspectives. These competencies identify skills necessary for students to have a well-rounded accounting education regardless of their career path, and they foster lifelong learning.

Functional competencies include:

1. Decision Modeling
2. Risk Analysis
3. Measurement
4. Reporting
5. Research
6. Technology

Personal competencies include:

1. Professional demeanor
2. Problem solving and decision making
3. Interaction
4. Leadership
5. Communication
6. Project management
7. Technology

Broad business perspective competencies include:

1. Strategic and critical thinking
2. Industry perspective
3. International perspective
4. Resource management
5. Legal and regulatory perspective
6. Marketing
7. Technology

The above competencies are interrelated and overlap. Obviously, the competencies cannot be developed in accounting courses alone. A strong business curriculum will provide the necessary background. A capstone course such as Administrative Decision-Making serves to bring together functional, personal and broad business perspective competencies. One cannot overlook the contributions to developing personal competencies by being involved in student organizations such as an Accounting Club.

Many accounting programs have responded over time to recommendations of the accounting profession, but accounting educators and practitioners need to work together to improve implementation of these efforts. For example, the providing of an internship allows the practitioner to take an active role in developing the necessary competencies. In order to make curriculum changes, an academic program must be analyzed from, at least, the bases of alumni responses and intern supervisors' evaluations. For this study, we survey accounting alumni on their perceptions of the importance of some of the competencies and attributes discussed above, as well as their preparation in these areas. In the next section, we present the summary and analysis of the recent alumni survey.

ALUMNI SURVEY

A sampling of accounting alumni consisted of 1,000 surveys sent out with 119 respondents. Table 1 shows the general background of the respondents. The emphasis of analysis will be on Table 2 (pages 51-52) related to the skills and abilities necessary to pursue a career in accounting. These skills and abilities relate to the functional, personal and broad business perspective competencies. The ranking of the skills by mean values includes:

1. oral communication skills
2. decision making ability
3. interpersonal skills
4. analytical thinking skills
5. ability to work independently
6. written communication
7. microcomputer skills
8. teamwork skills
9. technical skills

The results show the importance of oral communication skills but not as high a rating for written communication skills. The ranking of technical skills is surprising. However, we note that the rankings are for relative importance, and in an absolute sense all nine of the skill and ability categories were rated above the scale mid-point, indicating that respondents believed that all were important.

A statistical analysis for each skill or ability is shown in Table 3 (page 53). The analysis compares favorably with Table 2 rankings. However, the following should be noted:

1. **Total vs. Skill ratings**: Technical skills has the lowest standard deviation, which strongly indicates a general agreement upon that technical skills is generally unimportant related to other skills for a person to pursue a career in the accounting profession.

2. **Gender vs. Skill ratings**: From a gender perspective, males rank oral communication (1.64) low and interpersonal skills (1.75) and technical knowledge (2.29) and team work (2.16) low. Moreover, the deviations in the two lowest ranking items are lowest among the overall skill ratings, which indicates males generally think the technical knowledge and team work are relatively unimportant in career pursuit. Nevertheless, females tend to have a different view of skill ratings. Decision making ability (1.58) rates as the number one most important skill in career pursuit. In addition, its deviation (0.95) is the second lowest value of all. In addition, team work and technical skills rank the lowest of all for females.

3. **CPA and Non-CPA vs. Skill ratings:** CPAs rank analytical thinking skills (1.58) the most important skill in their careers, followed by oral communication skills (1.63). However, team work (2.29) was low in importance, as were computer and technology skills (2.13). In addition, the deviations among all skill ratings are relatively low. The value is from 0.88 to 1.16. Skills surveyed are generally lower rated among Non-CPAs. Generally, Non-CPAs tend to rank oral communication high (1.70) and technical knowledge (2.32) and team work (2.03) low. Further, the standard deviations among Non-CPAs' skill ratings is somewhat larger than those CPAs' skill ratings.

   When we look at this issue from the gender perspective, we note that male CPAs place more weight on the importance of oral communication and analytical thinking skills, which both have mean value of 1.48. Moreover, male CPAs rank the computer and technology skills the lowest (2.14) and team work the second lowest (2.05). This finding seems anomalous and may be unique to the limited sample responding. Female CPAs seem to have more conservative views on ranking skills. The most significant skills necessary to their careers are analytical thinking skills (1.71) and oral communication and decision making ability, which both ranked second of all (1.82). The least important skill rated was ability to work in a group (2.59). It is possible that there are different professional expectations and or career paths for female CPAs in comparison to their male counterparts.

4. **Partners with CPA, Public Accounting Firms and Accounting Profession in other Industries vs. Skill ratings**: There are some very interesting findings. Partners with a CPA all agree that interpersonal skills are the most important skills in their career pursuit. The standard deviation is 0. In addition, oral communication, written communication and decision making ability rank high (1.20) while their standard deviation (0.45) is significantly lower than the rest of the skill ratings.

   Accounting professionals in public accounting firms rank oral communication (1.96) as the most important skill. Team work ranks the lowest among skill ratings. In addition, technical knowledge has the lowest standard deviation (0.75), indicating that respondents generally agree with its relative unimportance (mean: 2.26).

   Accounting professionals in industry rank decision making ability (1.68) as the most important skill necessary to career pursuit. Technical knowledge (2.21) and team work (2.08) are ranked as the two least important skills necessary to career advancement. However, analytical thinking has the highest standard deviation (1.60), indicating that respondents generally disagree with the relative importance (mean:1.88).

5. **Big Five Accounting Firms vs. Skill ratings**: Accounting professionals in Big Five accounting firms (including Arthur Andersen) rank teamwork (1.38) as the most important skills necessary to their career, with oral communication (1.50) as the second, although its standard deviation indicates relative dispersal among ratings. Further, written communication (2.00) and technical skills (2.00) ranked lowest in career pursuit. A possible explanation is that the major public firms require auditors to work in teams, whereas small public accounting firms and other organizations may emphasize more individual assignments.

The summary results in Table 4 (page 54-55) of the intern supervisors' evaluations are presented as a way of assessing the overall effectiveness of accounting students' preparation for the profession. These results clearly show that our interns have the skills and abilities necessary to pursue a career in accounting. For example, oral and written communication skills evaluations both fall within the good to excellent range. Analytical thinking skills are covered under the criteria of knowledge of theory and application within the good to excellent range. The opportunity for decision making does not arise at the internship level. Interpersonal skills are addressed under the personal qualifications within the good to excellent range. The ability to work independently is addressed under initiative and interest in work; preparation and accuracy are addressed within the good to excellent range.

The microcomputer skills and technical skills are assumed to be applied in the evaluation criteria. Teamwork skills do not always apply to every internship situation. The category of growth potential indicates the interns will be able to develop within the profession.

In summary, the supervisors' evaluations indicate the success of our interns but give no indication of relative importance of each of the categories or criteria. The alumni survey helps with the emphasis within curriculum development.

IMPLICATIONS FOR ACCOUNTING CURRICULUM

The accounting profession and academia have collaborated to enhance accounting education primarily by expanding the educational requirements, in most states, to 150 credit hours necessary to either sit for the CPA examination or become certified. The rationale for increasing the educational requirements frequently focuses on greater opportunities, based on course selections, to develop non-technical skills.

The educational issues challenging the future of the accounting profession include (French, 2000):

1. Specification and development of the educational product.
2. Identification of key constituents and their needs.
3. Changing to a learning paradigm from a teaching paradigm.
4. Increased accountability to constituents for accomplishments of educational goals.
5. Establishing a strong partnership among secondary education, higher education and the accounting profession.
6. Changing role of accounting educators as managers of the educational processes.
7. Changing role of accounting students as active learners.
8. Changing role of accounting professionals as active participants in the recruitment of prospective students and delivery of accounting education.
9. Increased rewards to attract "the best and the brightest."
10. Lifelong learning opportunities for accounting professionals that focus on both technical and non-technical skills.

These issues indicate the breadth of thinking on accounting education in the profession, but go far beyond what we hope to accomplish with our research. However, with the issues specified we have tried to address the specification of the educational product and identification of key constituents and their needs.

Accounting educators and practitioners continue to debate and argue over preparation of entry-level accountants. Many reports characterize accounting education as static, not changing to meet professional needs and demands. Today's graduates do grasp basic technical knowledge; however, studies show they lack an understanding of new technology, communication skills, business ethics, business globalization, and multidisciplinary approaches to business decision-making (Simmons and Williams, 1996).

What students lack is a practical product of a dynamic, ever changing, more complex business environment. These business changes have placed new demands on accountants. There is always a lag between the profession and academia.

The accounting curriculum at Clarion University, in regard to courses listed in the catalog, has remained constant over the past 10 years. Actual offerings, however, have declined, especially the upper division accounting electives. Even though course offerings have declined, the necessary and important skills and abilities identified by the survey are still being achieved, according to the intern employers' evaluations. In general, the curriculum now offers fewer opportunities to broaden one's technical accounting knowledge beyond the core of Intermediate Accounting, Cost Accounting and Federal Taxes. This is being replaced, however, by courses involving a greater emphasis on oral and written communication as well as computer and technology skills. A consultant visited our department recently as part of our Five-Year Review. Based on our staffing, he recommended that some topics be existing required accounting courses such as Advanced Accounting.

The educational model must be modified to promote students' ability to learn real-world skills. Students need to spend more time engaged in activities that develop business skills and knowledge. The thrust is not to change course offerings but to work as a faculty team to modify course content and student activities within each upper-level, required accounting course. With declining enrollments, the opportunity for greater individual student attention could help to develop stronger skills in areas such as oral and written communications. The entire academic curriculum cannot be overlooked, as it is necessary for providing the background to develop the important skills and abilities. The accounting curriculum must use fundamental skills and abilities as a basis to build on higher level professional needs.

The accounting internship provides an opportunity for integrating skills and abilities. Any weaknesses in the curriculum possibly may be overcome by the practical experience. Also, the internship helps overcome the lag effect of necessary professional requirements being implemented on a timely basis by academia.

The AICPA pre-professional competency task force developed a framework that supports a paradigm shift from a content-driven to a skills-based curriculum. This is in response to a rapidly changing profession (Foster, Bolt-Lee, and Colson, 2002). The internship certainly solidifies a skills-based curriculum.

The task force leaves it to each accounting administrator to vary their program to best suit individual "environments, resources and missions" (Foster, Bolt-Lee, & Colson, 2002). Our rural environment, very limited staffing and mission of education of the masses mean we cannot use a "cookie-cutter" approach to developing our accounting curricula.

SUMMARY AND CONCLUSION

The results of our survey indicate an increasing importance for a variety of communication, interpersonal and integrative skill areas and a somewhat reduced importance for technical skills and, especially, technical accounting knowledge. While this finding is somewhat surprising, it may be that such skills are requisite to entry level positions, and while a satisfactory level is attained initially, further development in these areas is less important than other skills for career advancement. In addition, technical accounting knowledge is being de-emphasized in professional writings such as **What Corporate America Wants in Entry-Level Accountants** (IMA, 1994) and **Accounting Education: Charting the Course through a Perilous Future** (Albrecht & Sack, 2000). Technical accounting knowledge is increasingly being viewed as a resource that may be acquired on an "as needed" basis. We also need to recognize that, due to limitations of our response and sample size, it is possible these findings may simply be an anomaly limited to Clarion University.

The analysis of the internship supervisors' evaluations provides clear evidence on both the importance of the internship experience as a tool for integrating the skills and knowledge learned in the classroom and the effectiveness of our program in achieving that end. The accounting internship is arguably the important capstone experience necessary to complete a high quality accounting education.

Many studies have been undertaken that conclude that accounting curricula lacked relevance and that students' skills were not being developed adequately. The focus has shifted to competencies, or skills, rather than subject content or a body of knowledge because of the rapidly changing nature of the accounting profession. These competencies include functional, personal and broad business perspective.

Curriculum changes in an academic program must be analyzed from at least the bases of alumni responses and intern supervisors' evaluations. An alumni survey verifies the relative importance of various skills and abilities. The summary results show that our interns have the skills and abilities necessary to pursue a career in accounting. The educational model must be modified to promote students' ability to learn real-work skills and abilities.

REFERENCES

Accounting Education Change Commission (AECC). (1990, Fall). Objectives of education for accountants: Position statement number one. *Issues in Accounting Education,* 307-312.

Albrecht, W. S., & Sack, R. J. (2000). *Accounting Education: Charting the Course through a Perilous Future.* Accounting Education Series, Volume 16. Sarasota, FL: American Accounting Association.

Albrecht, W. S., & Sack, R. J. (2001). The Perilous Future of Accounting Education. *CPA Journal, 71(3),* 17.

American Accounting Association (AAA), Committee on the Future Structure, Content, and Scope of Accounting Education (The Bedford Committee). (1986, Spring). Future accounting education: Preparing for the expanding profession. *Issues in Accounting Education,* 168-195.

American Accounting Association (AAA). Report of the Changing Environment Committee. (1998). *The*

Future Viability of Accounting Education. Sarasota, FL: American Accounting Association.

Foster, S., Bolt-Lee, C., & Colson, R. (2002). New Competencies for Accounting Students. *CPA Journal, 72(1),* 68-71.

French, G. Richard. (2000). Educational Issues Challenging the Future of the Accounting Profession. *Ohio CPA Journal, 59(3)*, 69-73.

Institute of Management Accountants (IMA). (1994). *What Corporate America Wants in Entry-Level Accountants*. Executive summary (with the Financial Executives Institute). Montvale, NJ: Institute of Management Accountants.

Simmons, S. A., & Williams, A. W. (1996). What do accounting practitioners want? How do entry-level accountants measure up? *Arkansas Business and Economic Review, 29(1),* 1-9.

Thomas W. Oliver, Ph. D. University of Massachusetts is Professor of Accountancy Clarion University, Clarion, PA. His other research interests include financial reporting, capital markets and accounting education.

Charles J. Pineno, Ph.D. The Pennsylvania State University, is Professor of Accounting at the Byrd School of Business, Shenandoah University, Winchester, VA. His other research interests include management decision making, accounting education, and balanced scorecard systems.

Table 1
General Background Information

Gender Relationships:

Total population = 117

		Males	% of pop.	Females	% of pop.
# of respondents		55	47%	62	53%
Participation in the Internship Program		29	53%	45	73%
Graduate Degrees earned		12	22%	10	16%
Males = 11 Masters, 1 Doctoral; Females = 9 Masters, 1 Doctoral					
Recent salary range:	0 - 9,999	0	0%	1	2%
	10,000-19,999	2	4%	2	3%
	20,000-29,999	0	0%	9	15%
	30,000-39,000	11	20%	18	29%
	40,000-49,999	12	22%	9	15%
	50,000-59,999	6	11%	11	18%
	60,000-69,000	7	13%	4	6%
	70,000-79,000	2	4%	1	2%
	80,000-89,000	2	4%	0	0%
	90,000-99,000	1	2%	2	3%
	100,000+	11	20%	3	5%
Average recent salary		60,277		50,462	
Salary range of 1st job:	0-9,999	2	4%	2	3%
	10,000-19,999	22	40%	31	50%
	20,000-29,999	22	40%	17	27%
	30,000-39,000	8	15%	8	13%
Average salary of 1st job		21,667		20,345	
First job out of college in actg		47	85%	55	89%
# of employers since Clarion:	0	0	0%	1	2%
	1	17	31%	18	29%
	2	14	25%	13	21%
	3	7	13%	16	26%
	4	4	7%	4	6%
	5	5	9%	4	6%
	6	3	5%	3	5%
	7	0	0%	1	2%
	8	2	4%	0	0%
	9	0	0%	0	0%
	10+	2	4%	2	3%
Average # of employers since Clarion		3.2		2.8	
Taken CPA exam		31	56%	38	61%
Passed the CPA exam		19	35%	17	27%
Taken CMA exam		2	4%	2	3%
Passed the CMA exam		1	2%	1	2%

Table 2
Skills and Abilities

		Males	% of pop.	Females	% of pop.
1. Oral communication skills:	1	35	64%	34	55%
	2	12	22%	18	29%
	3	0	0%	6	10%
	4	5	9%	2	3%
	5	2	4%	2	3%
2. Written communication:	1	23	42%	30	48%
	2	16	29%	23	37%
	3	6	11%	3	5%
	4	5	9%	2	3%
	5	3	5%	3	5%
3. Interpersonal skills:	1	30	55%	32	52%
	2	16	29%	21	34%
	3	1	2%	4	6%
	4	4	7%	4	6%
	5	3	5%	1	2%
4. Microcomputer skills:	1	18	33%	29	47%
	2	18	33%	20	32%
	3	13	24%	8	13%
	4	3	5%	1	2%
	5	2	4%	3	5%
5. Decision making ability:	1	26	47%	38	61%
	2	19	35%	13	21%
	3	1	2%	4	6%
	4	4	7%	3	5%
	5	4	7%	1	2%
6. Ability to work independently:	1	23	42%	31	50%
	2	18	33%	19	31%
	3	6	11%	5	8%
	4	4	7%	2	3%
	5	3	5%	3	5%
7. Teamwork skills:	1	16	29%	23	37%
	2	21	38%	19	31%
	3	10	18%	12	19%
	4	6	11%	6	10%
	5	1	2%	1	2%
8. Technical skills:	1	12	22%	15	24%
	2	21	38%	29	47%
	3	15	27%	13	21%
	4	2	4%	2	3%
	5	3	5%	2	3%

		Males	% of pop.	Females	% of pop.
9. Analytical thinking skills:	1	30	55%	36	58%
	2	14	25%	13	21%
	3	4	7%	7	11%
	4	3	5%	2	3%
	5	3	5%	4	6%

Table 3: Skill Ratings by CPA and Gender

CPA = 38			Non-CPA = 81		
Skill Ratings	Mean	STD.	Skill Ratings	Mean	STD.
Oral	1.63	1.02	Oral	1.70	1.06
Written	1.95	1.16	Written	1.87	1.11
Interpersonal skills	1.68	1.04	Interpersonal skills	1.76	1.06
Computer skills	2.13	0.88	Computer skills	1.89	1.13
Decision making ability	1.71	1.04	Decision making ability	1.74	1.12
Ability to work independently	1.97	1.00	Ability to work independently	1.85	1.17
Team work	2.29	1.04	Team work	2.03	1.05
Technical knowledge	1.97	0.94	Technical knowledge	2.32	1.00
Analytical thinking	1.58	1.03	Analytical thinking	1.90	1.23

Male CPA = 21			Male Non-CPA = 35		
Skill Ratings	Mean	STD.	Skill Ratings	Mean	STD.
Oral	1.48	0.93	Oral	1.74	1.20
Written	1.81	1.12	Written	2.18	1.24
Interpersonal skills	1.62	1.07	Interpersonal skills	1.83	1.20
Computer skills	2.14	0.85	Computer skills	2.09	1.17
Decision making ability	1.62	1.07	Decision making ability	2.06	1.26
Ability to work independently	1.81	0.93	Ability to work independently	2.11	1.28
Team work	2.05	0.86	Team work	2.23	1.14
Technical knowledge	1.95	1.02	Technical knowledge	2.50	1.02
Analytical thinking	1.48	1.03	Analytical thinking	2.00	1.24

Female CPA = 17			Female Non-CPA = 46		
Skill Ratings	Mean	STD.	Skill Ratings	Mean	STD.
Oral	1.82	1.13	Oral	1.67	0.95
Written	2.12	1.22	Written	1.64	0.94
Interpersonal skills	1.76	1.03	Interpersonal skills	1.71	0.94
Computer skills	2.12	0.93	Computer skills	1.73	1.09
Decision making ability	1.82	1.01	Decision making ability	1.48	0.92
Ability to work independently	2.18	1.07	Ability to work independently	1.63	1.05
Team work	2.59	1.18	Team work	1.86	0.95
Technical knowledge	2.00	0.87	Technical knowledge	2.18	0.97
Analytical thinking	1.71	1.05	Analytical thinking	1.82	1.23

Table 4: Accounting Interns
Supervisors' Evaluation Reports
(Full-time & Part-time Internships)

No. of Students:	36	36	28	17	23	14	16	20	17
No. of Firms:	31	34	27	16	21	10	12	15	12
SPRING SEMESTER OF:	1993	1994	1995	1996	1997	1998	1999	2000	2001
A. PERSONAL QUALIFACTIONS Poise, manners, tact, cooperative spirit, appearance, professional bearing, attitude, punctuality 4= Outstanding; makes excellent impression 3= Creates good impression in most respects 2= Acceptable; has one or two short-comings (specify) 1= Creates unfavorable impression in several respects (specify)	3.58	3.57	3.75	3.77	3.73	3.78	3.87	3.60	3.76
B. KNOWLEDGE OF THEORY 4= Has a thorough understanding 3= Has adequate understanding 2= Weak in a few areas, otherwise satisfactory (indicate weak areas) 1= Generally weak (comment on major areas)	3.19	3.34	3.50	3.35	3.39	3.50	3.50	3.725	3.62
C. INITIATIVE AND INTEREST IN WORK 4= Anxious to accept responsibility; has real drive 3= Good worker, produces good results 2= Carries out assignments but little beyond 1= Must be supervised closely to see that he does assigned work	3.47	3.57	3.73	3.65	3.65	3.85	3.75	3.625	3.65
D. ABILITY TO FOLLOW INSTRUCTIONS 4= Follows instructions well, implements them with constructive ideas 3= Follows instructions, asks questions when in doubt 2= Sometimes forgets instructions, hesitates to ask questions 1= Frequently disregards instructions	3.19	3.46	3.58	3.29	3.43	3.64	3.50	3.75	3.53

SPRING SEMESTER OF:	1993	1994	1995	1996	1997	1998	1999	2000	2001
E. APPLICATION 4= Exceptionally efficient in accomplishing tasks 3= Works fast; seldom gets bogged down 2= Keeps working, gets done on time with a little prodding 1= A slow worker; seldom gets done on time	2.98	3.28	3.46	3.53	3.21	3.64	3.37	3.60	3.41
F. PREPARATION AND ACCURACY 4= Interested in improving work; presents information and conclusions in orderly and legible manner 3= Papers well prepared; generally accurate, complete and legible 2= Generally satisfactory; sometimes omits essential information 1= Papers frequently incomplete, inaccurate, poorly arranged and illegible	3.37	3.47	3.75	3.71	3.52	3.71	3.75	3.725	3.60
G. ORAL EXPRESSION 4= Expresses himself/herself very well 3=Usually expresses himself/herself well 2= Generally adequate; sometimes hesitant or unsure of himself/herself 1= Has not yet developed ability to express himself/herself adequately	3.38	3.36	3.78	3.47	3.52	3.57	3.56	3.55	3.59
H. WRITTEN EXPRESSION 4= Expresses himself/herself very well; written material requires little or no editing 3= Usually expresses himself/herself well 2= Adequate; sometimes has difficulty in expression 1= Has not developed ability to express himself/herself adequately	3.29	3.42	3.57	3.64	3.38	3.38	3.80	3.50	3.54
I. GROWTH POTENTIAL 4= Developing rapidly; excellent potential 3= Has growth potential; is developing satisfactorily 2= Approaching limit 1= Not capable of performing presently assigned work	3.48	3.50	3.62	3.88	3.69	3.78	3.75	3.725	3.91

THE PREDICTIVE POWER OF THE SURVEY OF PROFESSIONAL FORECASTERS' PROBABILITY FORECASTS

J. George Wang, College of Staten Island

ABSTRACT

The probability forecast of a decline in real GDP from the Survey of Professional Forecasters (SPF) has long been used as a predictor of the cyclical movement of the economy by various users in the public and private sectors. However, little attention has been paid to its predictive power and forecasting performance in the literature. In this paper, the PT Predictive Power Test and the Kuipers Score are applied to assess the usefulness of the SPF's probability of decline in real GDP as an indicator of the future path of the economy.

INTRODUCTION

The cyclical movement in real GDP has long been the focus of business cycle researchers and business practitioners. As witnessed for past decades, the economy was unexceptionally impacted by the cyclical movement of real GDP for every downturn and upturn, and an early detection of the phase change in the business cycle could provide enormous value for corporations, individuals, and government policy makers. For this purpose, professional forecasts from a variety of forecasting entities using various forecasting techniques were conducted with the intention to supply the government and corporate decision makers with some reliable and timely guidance for the future path of the economy. Among them, the probability forecast of decline in the GDP from the Survey of Professional Forecasters (SPF) is one that is being constantly monitored and frequently used by various end-users in both public and private sectors.

Given the critical role of the prediction of the future movement of real GDP in influencing business decisions, one important issue in business cycle research is the quality of the forecasts. While a high quality forecast can provide its end users with a useful leading indicator for their business references, a low quality forecast could be a "misleading" indicator in terms of the direction, timing, and magnitude of the future changes in the economy. Therefore, any professional forecast, including the probability forecasts of decline in real GDP, without associated evaluations, should be considered a mission incomplete and used with extra caution.

Evaluation of the forecasts can be performed in different ways depending upon the type of forecast. In addition to the traditional point forecast, interval forecast, density forecast, probability forecast, direction/event forecast, and their associated evaluation methodologies are well developed in the literature. Christoffersen (1998) developed a general condition-efficiency criterion for evaluating the interval prediction. A likelihood ratio test of conditional coverage was proposed with an application to daily exchange rates. Diebold, Gunther and Tay (1998) suggested an evaluation method in a decision-theoretical framework for the density forecast. Pesaran and Timmermann (1992) proposed a non-parametric test statistic to evaluate the predictive power of the direction/event forecasts with the null hypothesis of independence between the forecast and the occurrence of the event. In addition, Granger and Pesaran (2000) established the linkage between a measure of forecast accuracy known as the Kuipers Score and the market timing test in finance with an empirical application to the problem of stock market predictability, for example.

The purpose of this paper is to evaluate the SPF by assessing the predictive power and the performance score of the probability of decline in the real GDP with different forecasting horizons. In contrast to the importance of the SPF probability forecasts for GDP decline in the business decisions, little attention has been paid to the evaluation of its predictive power and performance score in the literature. It is hoped that the evaluation of the performance of the SPF probability for a decline in the real GDP in this paper will provide its end-users with a needed assessment for its usefulness as a predictor of the cyclical movement of real GDP.

The structure of the paper is as follows. Section II presents the descriptions and empirical data of the SPF probability forecasts. Section III assesses the predictive power of the SPF forecasts. Section IV analyzes the forecast score in terms of missing signals and false alarms. Finally, Section V ends the article with some concluding remarks.

THE SPF PROBABILITY FORECASTS ON DECLINE IN REAL GDP

As one of the oldest business surveys in the US since 1968, the American Statistical Association and the National Bureau of Economic Research (ASA/NBER) routinely conducted quarterly surveys for the forecasts of the future economy from professional forecasters. The questionnaires are mailed out when the forecasters typically review and update their predictions, and the responses are received by the middle of the second month of the quarter. The number of the responses to the questionnaires usually ranges from about 20 to 150. The survey was commonly referred to as the ASA-NBER Survey in the previous literature, and the name was changed to the Survey of Professional Forecasters (SPF) when the Federal Reserve Bank of Philadelphia took over responsibility for the survey in June 1990.

The SPF generally covers the forecasting horizons of the current quarter and up to the subsequent four quarters. In addition to the probability of a decline in the real GDP, the variables to be predicted also include other GDP-related measures, inflation, the unemployment rate, and other important macro-economic variables that are closely watched by government and business decision makers, media, and the general public.

The graphs of the mean probability of decline in real GDP in the current and the following quarters from Quarter 4 1968 to Quarter 2 2004 are depicted in Figures 1-5 (pages 62-64). The lines in each figure display the probability of decline in the real GDP in different quarters, as the professional forecasters made the predictions over time, and the real time real GDP growth rate, respectively. For the real time real GDP growth, the version of the July Revision is used. It is well known that the July Revision of real GDP release is a relatively complete one in each year. It is released in each "current" year, so that it is real time data compared with the final revision. Meanwhile, it captures all the major consecutively negative changes of real GDP which are sometimes missed by the quarterly (or 30-Day) release (Q1 and Q2 2001, for example).

From Figures 1-5, several notable phenomena can be observed. First, the mean probabilities generated by the professional forecasters fluctuate over time in a certain pattern. The value of the mean probability varies from as high as the 80% range to as low as less than 5%. Second, the fluctuation in the mean probability seems coincident with the fluctuations in real GDP growth. That is, around the time with the negative growth rate of the real GDP or the recessionary periods, the probabilities suddenly rise up; and in the time associated with the positive growth rate or the expansionary periods, they remain relatively low. Third, for the different forecasting horizons, the sudden increases or decreases in the probabilities either precede or follow the cyclical movement of real GDP with different time leads or lags. Finally, the high end of the mean probability tends to decrease as the forecasting horizon increases. As shown in the Figures, the high end probability decreases from above 80% for the current quarter to the 70% range for the one-quarter-ahead, to the 50% range for the two-quarters-ahead, and to the 30% range for the three- and four-quarters-ahead forecasting horizons. Consequently, the correlation between the SPF probability and the GDP growth decreases from high to low as the forecasting horizon increases (0.63, 0.47, 0.28, 0.19, and 0.09 for the current quarter and the following four quarter forecasting horizon, respectively).

All these observations indicate that the probability forecasts for the decline in the real GDP contain tremendous information about the phase changes of the real GDP. Consequently, some assessments need to be conducted for evaluating their forecasting ability.

PREDICTIVE POWER OF THE SPF FORECASTS

The predictive power of the SPF probability forecasts is examined in this section. Given the binary nature of the event, the Pesaran and Timmermann (PT) test (Pesaran & Timmermann, 1992) is used to assess the predictive power of the SPF probability forecasts in their abilities to predict the future path of the real GDP growth.

The PT test was designed to test the prediction of directional changes such as an occurrence or a non-occurrence of an event (for example, the decline in real GDP or non-decline in real GDP for the current quarter or for future quarters). However, the SPF forecasts are expressed as the probabilities that the real GDP will decline in different forecasting horizons. Therefore, the SPF probability needs to be first translated into an event variable to perform the PT test. One common method of translation is the traditional naïve approach. That is, a value of 1 will be assigned to a dummy variable if the forecasting probability for the occurrence of the event is above 50%; otherwise a value 0 will be assigned.

For the PT test, the forecast evaluation is conducted by calculating the difference between the proportion of the times that the event is predicted correctly and the mean of the underlying binomial distribution (the theoretical portion of the times of the occurrence of the event) under the null hypothesis of independence between the forecast and the occurrence of the event in the 2x2 case. The test statistic is structured as follows:

$$S_n = (\hat{P} - \hat{P}_*)/(Var(\hat{P}) - Var(\hat{P}_*))^{1/2} \qquad (1)$$

where $\hat{P}$ is the proportion of the times that the event is predicted correctly,

$\hat{P}_* = P_y P_x + (1-P_y)(1-P_x)$, is the mean of the binomial distribution under the null hypothesis,

P_y and P_x are the probability of the occurrence of the event and the forecasted probability of the occurrence of the event, respectively. When the theoretic P_y and P_x are unknown, P_y and P_x can be efficiently estimated by

$$\hat{P}_y = \sum_{t=1}^{N} Y_t / N$$

$$\hat{P}_x = \sum_{t-1}^{N} X_t / N$$

respectively, under the null; where Y_t is the occurrence of the event, and X_t is the prediction of the occurrence of the event.

$Var(\hat{P}) = \hat{P}(1-\hat{P}_*)/N$, is the variance of $\hat{P}$,

$$Var(\hat{P}_*) = (2\hat{P}_y - 1)^2 \hat{P}_x(1-\hat{P}_x)/N + (2\hat{P}_x - 1)^2 \hat{P}_y(1-\hat{P}_y)/N + 4\hat{P}_y\hat{P}_x(1-\hat{P}_y)(1-\hat{P}_x)/N^2$$

is the variance of $\hat{P}_*$.

Under the null hypothesis of independence between the SPF probability forecasts and the occurrence of the event, the difference between the proportion of the correct forecasts made by SPF ($\hat{P}$) and the proportion of the occurrence of the event ($\hat{P}_*$) should be insignificant as measured by the test statistic S_n in (1), which follows $N(0,1)$ under the null. Conversely, if the test fails at any acceptable level of significance, a dependent relationship between the two series will be considered. Therefore, the existence of the predictive power of the forecast on the occurrence of the event will be supported by the data.

Given the directional nature of the PT test and the probability expression of the SPF forecast as discussed above, one can use the naïve approach with 50% as the threshold of the directional indicator for the SPF forecast series. The SPF probability of decline in real GDP above 50% will be considered a direction change (down); otherwise, it will be considered a non-down prediction. The test results are displayed in Table 1 (page 65).

As it turns out, the test results for both current and next quarter forecasts (two-quarter-ahead and above forecasts are not applicable, because none or too few forecast probabilities are above 50%) uniformly reject the independence hypothesis with any commonly used acceptable level, indicating strongly the existence of the predictable relationship between the SPF forecasts and the actual decline in real GDP. In other words, the SPF forecast is not a groundless predictor for the occurrence of the decline in real GDP in the current and the following quarter. Instead, the SPF contains useful information about the target being predicted, and, thus, is an important source for watching the phase changes of real GDP cyclical movements.

KUIPERS SCORE OF FORECASTING PERFORMANCE

The performance of the SPF probability forecasts can also be evaluated in terms of the missing signals and the false alarms. As in any probability forecast, the trade-off between the missing signals and the false alarms always exists. In general, a decision rule based on the forecast probability that tends to decrease the missing signals or increase the "hit rate" will tend to increase the false alarms. So the balance of these "Type I" and "Type II" error is an important issue for the evaluation of the forecasts, and is usually determined by the loss function associated with the different decisions whenever it is available. In this regard, Kuipers Performance Index or Kuipers Score

(Granger & Pesaran, 2000) with the contingency matrix is a good measure to be used. It can be applied to the event variable that is translated from the SPF with the naïve approach as discussed above. The Kuipers Score (KS) was originally proposed by Pierce (1884) and was used widely in evaluating forecasting performance in meteorology. More specifically, KS is defined as the difference between the "hit rate" (H) and the "false alarm rate" (F) as follows:

$$KS = H - F \tag{2}$$

Where $H = T_{by} / (T_{by} + T_{bn})$
$F = T_{gy} / (T_{gy} + T_{gn})$

T_{by} is the number of times that the event occurred (the subscript "*b*" stands for the bad thing (event) happened) and the forecaster predicted it correctly ("*y*" for "yes" answer given by forecaster). T_{bn} is the number of times that the event occurred, but the forecaster failed to predict it (with "no" answer). So the ratio *H*, then, is the "hit ratio" that measures the portion of the times that the forecasters predict correctly when the event occurred.

Similarly, T_{gy} is the number of times that the event did not occur, but the forecasters mistakenly predicted it. T_{gn} is the number of times that the event did not occur and the forecasters correctly said no. So the ratio *F* is the "false alarm ratio" that measures the portion of times that the forecasters generated false signals for the occurrence of the event when it actually did not happen. By definition, the total number of the observations (*T*) is the sum of all *T's*. That is:

$$T = T_{by} + T_{bn} + T_{gy} + T_{gn}$$

Naturally, the higher the value of KS with a positive score, the better the forecasting performance with fewer missed events and fewer false alarms. In other words, the higher the KS, the higher the hit rate with a relatively low false alarm rate. Conversely, a lower KS, or even a negative KS will indicate a higher false alarm rate relative to the hit rate. Obviously, if a forecaster always predicts the occurrence of the event systematically, then the KS will be equal to 0 with 100% hit rate and 100% false alarm rate. Similarly, if a forecaster always predicts the non-occurrence of the event systematically, then, the KS will also be equal to 0, but with 0% false alarm rate and 0% hit rate.

The Contingency Matrix and the KS using the naïve approach with different forecasting horizons are displayed in Tables 2 and 3. The numbers in the first, second, third and fourth quadrants correspond to $T_{by}, T_{gy}, T_{gn}, T_{bn}$, for each forecasting horizon, respectively.

As Tables 2 and 3 (page 65)show, with the naïve approach, SPF generated a low false alarm rate less than 5% (F = 6/(6+117) = 0.048 for the current quarter, for example) for all forecasting horizons. The hit rate reaches its high end at 65% (H = 13/(13+7)) in the current quarter, but deteriorates to 30% (H = 6/(6+14)) for the one-quarter-ahead, and further goes down to 5% (H = 1/(1+19)) for the two-quarter-ahead as the forecasting horizon increases. Consequently, the KS follows a similar pattern. That is, the value of KS reaches its high end at 0.6012 (KS = 0.65 - 0.0488) in the current quarter, then goes down to 0.2593 (KS = 0.3 – 0.0407) in the one-quarter-ahead, and further down to about 2% (KS = 0.05 – 0.0325) for the longer forecasting horizon.

CONCLUDING REMARKS

In summary, the PT predictive power test and the Kuipers Score are used in this study to evaluate the forecast performance of the SPF probability forecasts for the decline in real GDP with different forecasting horizons. The conclusions of the paper are as follows.

First, the SPF probability forecasts for the decline in real GDP in the current quarter and the near future contain tremendous amounts of information about the regime switching of the cyclical movement in real GDP. They are indubitably important information sources for the possible signals of cyclical phase changes.

Second, the PT predictive power test reveals strong evidence for the existence of a dependent relationship between the forecasts and the event being forecasted, at least, in the shorter run, as indicated by the rejection of the null hypothesis of independence. That means the SPF probability forecasts are the predictions for the near future real GDP movement with the required predictive power.

Third, similar to the issue of the trade-off between the "Type I" and "Type II" errors, it is desirable that a higher hit rate is achieved with a lower false alarm rate as indicated by a higher Kuipers Score. As reported in this study, the current quarter forecast performs best with the highest Kuipers Score. Meanwhile, the performance of the SPF probabilities deteriorates as the forecasting horizon increases.

Finally, it should be noted that the SPF probability forecasts for the decline in real GDP, especially for the longer forecasting horizons, seem conservative. The probabilities of decline in the real GDP assigned by the forecasters are relatively low, even when the real GDP decline was almost around the corner or had already occurred. Therefore, the SPF probability for the decline in real GDP does contain correct directional information for real GDP, but may need to be used with some "adjustment" or amplifications for its magnitude. But that will be the issue of further research.

REFERENCES

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4), 841-862.

Diebold, F. X. Gunther, T. A., & Tay, A. S. (1998). Evaluating density forecasting with application to financial risk management. *International Economic Review*, 39(4), 863-883.

Granger, C. W. J., & Pesaran, M. H. (2000). Economic and statistical measures of forecast accuracy. *Journal of Forecasting*, 19, 537-560.

Peirce, C. S. (1884). The numerical measure of the success of prediciton. *Science*, 4, 453-454.

Pesaran, M. H., & Timmermann, A. (1992). A simple nonparametric test of predictive performance. *Journal of Business & Economic Statistics*, 10(4), 461-465.

The earlier version of this paper was presented at the 2004 Annual Conference of Association of Pennsylvania University Business and Economics Faculty at State College, Pennsylvania, on October 1, 2004.

J. George Wang, SUNY at Albany, is assistant professor of finance, chair of the finance area, business department, at College of Staten Island, CUNY, teaching advanced corporate finance, international finance, and international business. His current research interests are business cycle forecasting, and evaluation of probability forecasts, survey data.

Fig. 1: Probability of Decline in Real GDP in the Current Quarter

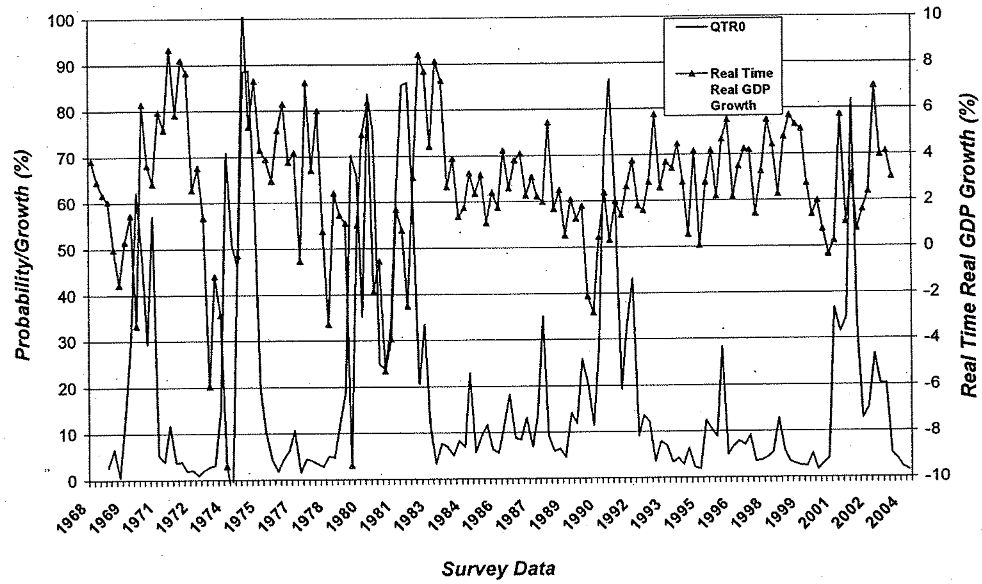


Fig. 2: Probability of Decline in Real GDP in the Following Quarter

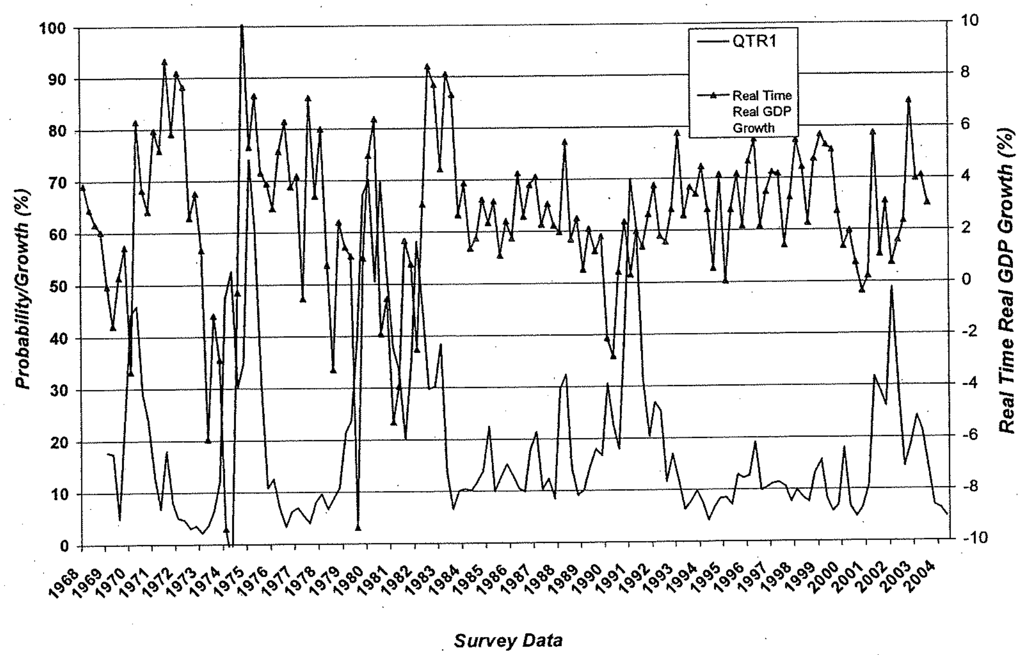


Fig. 3: Probability of Decline in Real GDP in Second Following Quarter

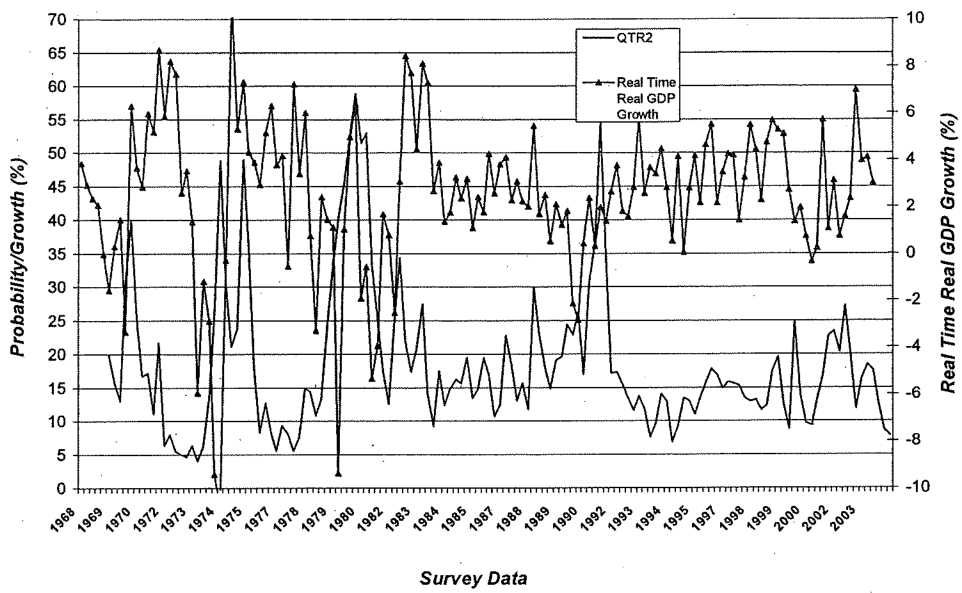


Fig. 4: Probability of Decline in Real GDP in Third Following Quarter

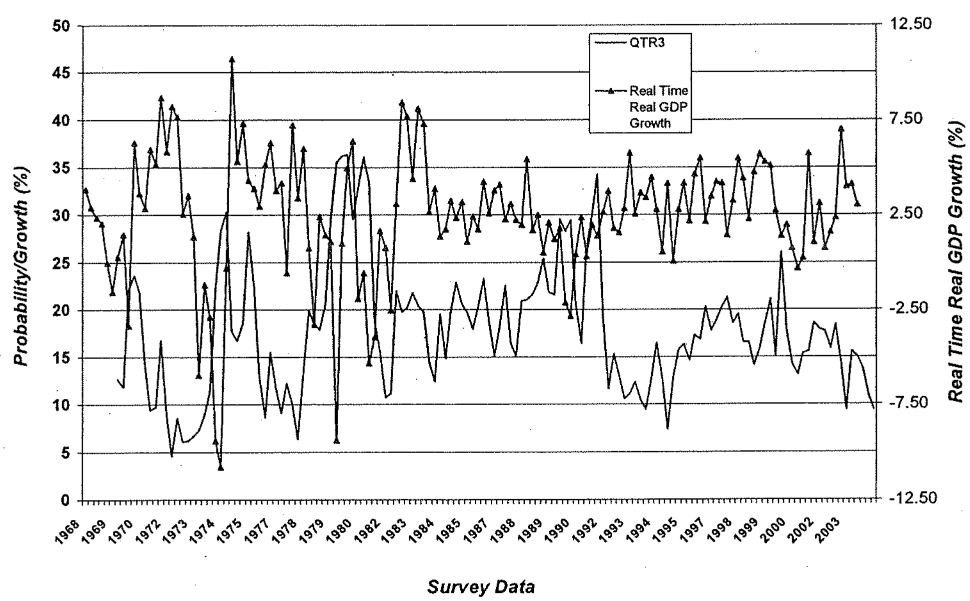


Fig. 5: Probability of Decline in Real GDP in Fourth Following Quarter

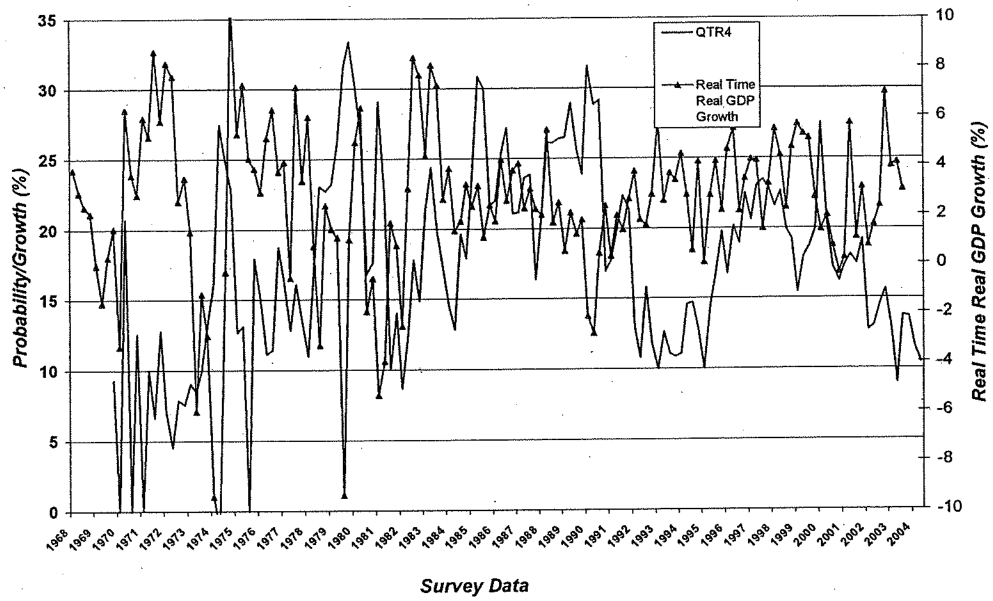


Table 1: PT Predictive Power Test

Horizon	PT Test
Q0	-34.33
Q1	-49.52
Q2	N/A
Q3	N/A
Q4	N/A

Table 2: Contingency Matrix

		Realization	
Horizon	Forecasts	Bad (Zt = 1)	Good (Zt = 0)
Q0	Yes	13	6
Q1	Yes	6	5
Q2	Yes	1	4
Q3	Yes	0	0
Q4	Yes	0	0
Q0	No	7	117
Q1	No	14	118
Q2	No	19	119
Q3	No	20	123
Q4	No	19	120

Note: (1) Yes (Zt = 1) represents the occurrence of the event
(2) No (Zt = 0) represents the non-occurrence of the event
(3) Total observations are 143 quarters from Q4 1968 to Q2 2004.

Table 3: Kuipers Score

Q0	Q1	Q2	Q3	Q4
0.6012	0.2593	0.0175	0.0000	0.0000

LINKAGES BETWEEN DJIA, S&P 500, AND NASDAQ COMPOSITE DAILY INDEX RETURNS IN BULL AND BEAR MARKETS

Ilhan Meric, Rider University
Mitchell Ratner, Rider University
Gulser Meric, Rowan University

ABSTRACT

This paper studies the co-movements of the DJIA, S&P 500, and NASDAQ Composite indexes during the July 17, 1998-March 24, 2000 and March 17, 2003-January 26, 2004 bull markets and the March 24, 2000-October 9, 2002 bear market. The findings indicate that the NASDAQ Composite Index has the most volatile daily returns and the DJIA Index has the least volatile daily returns in both bull and bear markets. The correlation between the returns of the three indexes has considerable volatility over time. The Granger-causality test results indicate that the past daily returns of each index can predict its own future daily returns in both bull and bear markets, i.e., none of the three daily index returns follow a random walk. The past daily returns of the NASDAQ Composite Index can predict the future daily returns of the other two indexes (i.e., the NASDAQ Composite index leads the DJIA and S&P 500 indexes) in a bull market. The past daily returns of the DJIA and S&P 500 indexes can predict the future daily returns of the NASDAQ Composite Index (i.e., the DJIA and S&P 500 indexes lead the NASDAQ Composite index) in a bear market. Investors can earn above-normal profits by following the signals from the movements of the leading stock market index in bull and bear markets.

INTRODUCTION

Studying the linkages between the U.S. stock market index and the stock market indexes of other countries has been one of the most popular research topics in finance (see Aggarwal & Kyaw, 2005; Hilliard, 1979; Meric & Meric, 1989; Philippatos, Christofi, & Christofi, 1983). However, the co-movements of U.S. domestic stock market indexes have not received sufficient attention. In this paper, we study the linkages between the daily returns of the Dow-Jones Industrial Average (DJIA), Standard and Poor's 500 (S&P 500), and NASDAQ Composite indexes with the rolling correlation analysis and Granger causality techniques. The DJIA, S&P 500, and NASDAQ Composite indexes are the most important U.S. stock market indexes carefully watched by all investors every day. For portfolio diversification purposes, studying the time-varying correlation and the lead/lag relations between these indexes would be of great interest to investors. Recent studies indicate that the co-movements of stock market indexes can change significantly from a bull market to a bear market (see Meric, Coopersmith, Wise, & Meric, 2002). In this paper, we study the co-movements of the DJIA, S&P 500, and NASDAQ Composite indexes in bull and bear-markets separately and compare the results.

DAILY RETURNS VOLATILITY

During the March 24, 2000-October 9, 2002 period, the U.S. stock market experienced one of the worst bear markets in its history. We study the co-movements of the DJIA, S&P 500, and NASDAQ Composite daily index returns during this bear-market period. To compare the results, we also study the co-movements of the returns during a bull-market period of equal length. The March 17, 2003-January 26, 2004 period was the most recent bull market in U.S. economy. The period since January 26, 2004 has been neither a bull market nor a bear market with stock market indexes moving temporarily in either direction with a high volatility mainly affected by worldwide events. The March 24, 2000-October 9, 2002 bear-market was preceded by a bull market. To have a bull-market period of equal length in the analysis with the March 24, 2000-October 9, 2002 bear-market period, the July 27, 1998-March 24, 2000 bull-market period was also added to the March 17, 2003-January 26, 2004 bull-market period. The average return statistics indicate that the July 27, 1998-March 24, 2000 bull market was a stronger bull market compared with the March 17, 2003-January 26, 2004 bull market. This gives us the opportunity to compare the results for these two bull markets of different strength as well.

The closing daily stock market index levels for the DJIA, S&P 500, and NASDAQ Composite indexes are downloaded from the Yahoo, Inc., web site. The daily index return figures are computed as the natural log difference in the indexes, ln ($I_{i,t}/I_{i,t-1}$). The daily returns of the three indexes for the July 27, 1998-March 24, 2000 and March 17, 2003-January 26, 2004 bull-market periods and the March 24, 2000-October 9, 2002 bear-market period are shown in Figure 1 (page 76). The graphs indicate that the daily returns of the NASDAQ Composite Index are considerably more volatile compared with the daily returns of the other two indexes both in the bull market and in the bear market.

The average daily returns of the three indexes and the standard deviations of the daily returns are presented in Panel A of Table 1 (page 72). The statistics in the table indicate that the NASDAQ Composite Index provides the highest average daily return during the total period (0.001570, in decimals). The S&P 500 Index provides the second highest average daily return (0.000831, in decimals), and the DJIA provides the lowest average daily return (0.000633, in decimals). The standard deviation figures, also measured in decimals, show that the NASDAQ Composite Index is the most volatile of the three indexes (0.023103). The S&P 500 Index is the second most volatile index (0.013425) and the DJIA is the least volatile index (0.012823). The returns of the three indexes appear to conform with the basic investment principle: the higher the volatility risk, the higher the return.

The return and standard deviation statistics indicate that the NASDAQ Index has the highest (0.001570/0.023103=0.068), the S&P 500 Index has the second highest (0.000831/0.013425=0.0619), and the DJIA Index has the lowest (0.000633/0.012823=0.0494) average daily return per unit of volatility risk in the total period.

The NASDAQ Index provides the highest average daily return, the S&P 500 Index provides the second highest average daily return, and the DJIA Index provides the lowest average daily return in both bull markets. The standard deviation statistics indicate that the NASDAQ Composite Index is the most volatile of the three indexes, the S&P 500 Index is the second most volatile index, and the DJIA Index is the least volatile index in both bull markets.

All three indexes have negative average daily returns in the March 24, 2000-October 9, 2002 bear market. The highest average daily loss is in the NASDAQ Composite Index, the second highest average daily loss is in the S&P 500 Index, and the lowest average daily loss is in the DJIA Index. All three indexes are considerably more volatile in the bear market than they are in the two bull markets. The NASDAQ Composite Index is the most volatile index, the S&P 500 Index is the second most volatile index, and the DJIA Index is the least volatile index in the bear market.

The comparisons in Panel B of Table 1 imply that the mean return of the NASDAQ index tends to be significantly higher than the mean returns of the other two indexes in a strong bull market, e.g, the July 17, 1998-March 24, 2000 bull market. The mean return of the NASDAQ index tends to be significantly lower than the returns of the other two indexes in a bear market, e.g., the March 24, 2000-October 9, 2002 bear market. Investors can earn above-normal profits by investing in the NASDAQ index in a strong bull market and by avoiding investing in the NASDAQ index in a bear market. However, in the long run covering equal lengths of bull and bear markets, returns earned from the NASDAQ index would not be significantly different from the returns earned from the other two indexes.

ROLLING CORRELATION ANALYSIS

Period correlation coefficients are generally used to determine the portfolio diversification benefit of index investments. However, recent studies demonstrate that the correlation between stock market index returns can be quite volatile over time (see Meric et al., 2002; Solnik, Boucrelle, & Le Fur, 1996). In this segment of the study, we use the rolling correlation analysis technique to study the time-varying correlation between the DJIA, S&P 500, and NASDAQ Composite daily index returns.

Starting with the first month, we computed the monthly rolling correlation coefficients among the three index returns by rolling the sample period ahead one trading day at a time for the entire July 27, 1998-January 26, 2004 period. Specifically, the latest daily observation is added while the earliest observation is deleted. A total of 1,245 rolling correlation coefficients are computed for each pair of indexes. The time-varying correlation graphs among the daily returns of the three indexes are presented in Figure 2 (page 77). The rolling correlation results for the bull and bear markets are marked on the graphs for comparison.

The graphs indicate that the correlation is very volatile between the daily returns of the three indexes both in the two bull markets and in the bear market. The correlation is generally quite high between the returns of the three indexes (please see Table 2 (page 73) for the actual correlation figures). However, the correlation becomes even negative between the DJIA Index returns and the NASDAQ Composite Index returns and between the NASDAQ Composite Index returns and the S&P 500 Index returns towards the end of the July 7, 1998-March 24, 2000 bull-market period, just before the beginning of the March 24, 2000-October 9, 2002 bear-market period.

The average rolling correlation coefficients between the daily returns of the indexes and the standard deviation and the coefficient of variation figures showing the volatility of the correlation coefficients during the two bull markets and the bear-market are presented in Table 2. The correlation among the three indexes is generally high in both bull and bear markets. All correlation coefficients are statistically significant at the one-percent significance level.

The average correlation coefficients for the total period indicate that there is a generally higher correlation between the DJIA Index returns and the S&P 500 Index returns and between the NASDAQ Composite Index returns and the S&P 500 Index returns than between the DJIA Index returns and the NASDAQ Composite Index returns. The standard deviation and the coefficient of variation figures show that the most volatile correlation is between the DJIA Index returns and the NASDAQ Composite Index returns, and the least volatile correlation is between the DJIA Index returns and the S&P 500 Index returns.

The statistics indicate that there is a higher correlation between the DJIA Index returns and the NASDAQ Composite Index returns in the March 17, 2003-January 26, 2004 bull-market period than in the previous two periods. The correlation is more volatile between these two indexes in the July 17, 1998-March 24, 2000 bull-market period compared with the other two more recent periods. The correlation between the DJIA Index returns and the S&P 500 Index returns is slightly higher in the March 17, 2003-January 26, 2004 bull-market period than in the previous two periods. The correlation is slightly more volatile between the returns of these two indexes in the March 24, 2000-October 9, 2002 bear-market period compared with the two bull-market periods. The correlation is higher and it is more volatile between the NASDAQ Composite Index returns and the S&P 500 Index returns in the July 17, 1998-March 24, 2000 bull-market period than in the two more recent periods.

The portfolio theory argues that the lower the correlation between investments, the greater the portfolio diversification benefit. The above results indicate that the DJIA and NASDAQ Composite indexes have the lowest correlation. Therefore, investment in these two indexes would provide the greatest portfolio diversification benefit, particularly in a strong bull market.

GRANGER-CAUSALITY TESTS

In several recent studies, the Granger-causality technique is employed to determine if the past returns of some national stock market indexes can be used to predict the future returns of some other national stock market indexes (see Meric et al., 2002; Ratner & Leal, 1996). An independent variable *X* Granger-causes changes in dependent variable *Y*, if *Y* can be better forecast with past values of *X* and *Y*, than just with past values of *Y* alone. The causality in the Granger sense does not imply a cause and effect relationship, but one of predictability. A detailed discussion of the Granger-causality technique is provided in Enders (1995).

Granger-causality is a useful time-series analysis technique to use when studying the lead/lag linkages between time-series data. It enables the researcher to study the predictive power of each time-series index of other time-series indexes. The technique also provides an opportunity to test if the past values of a given time-series index can be used to predict its own future values. If the time-series data used in the analysis are stock market index returns, it is a valuable opportunity to test the weak-form efficiency of the stock market.

We use the Granger-causality technique to study if the past daily returns of the DJIA, S&P 500, and NASDAQ Composite indexes can be used to predict each other's and its own future daily returns in bull and bear markets. The test statistics for the combined July 17, 1998-March 24, 2000 and March 17, 2003-January 26, 2004 bull-market periods are presented in Table 3 (page 74).

The test statistics indicate that the past NASDAQ Composite daily index returns can predict both the future DJIA daily index returns and the future S&P 500 daily index returns in a bull market at the one-percent level of significance. However, the past daily returns of neither of the other two indexes can predict the future daily returns of the NASDAQ Composite Index in a bull market. The past daily returns of the DJIA and S&P 500 indexes cannot predict the future returns of one another in a bull market at the conventional five-percent level of significance. The F statistics imply that the daily returns of none of the three indexes follow a random walk in bull markets, i.e., the past daily returns of each index can be used to predict its own future daily returns at the one-percent level of significance.

The Granger-causality test statistics for the March 24, 2000-October 9, 2002 bear-market period are presented in Table 4 (page 75). Unlike the case of the bull market, the past daily returns of the NASDAQ Composite Index are not a good predictor of the future daily returns of the other two indexes in the bear market at the conventional five-percent level of significance. The future daily returns of neither the DJIA Index nor the S&P 500 Index can be predicted by the past daily returns of the other two indexes. However, the test statistics indicate that the past daily index returns of the DJIA Index and the S&P 500 Index can predict the future daily index returns of the NASDAQ Composite Index at one-percent and five-percent significance levels, respectively. Again, as in the tests for the bull market, the past daily returns of each index can predict its own future daily returns at the one-percent level of significance (i.e., the daily returns of the indexes do not follow a random walk) in the bear market.

The Granger-causality test results in Table 3 indicate that the NASDAQ Composite Index has a leading role and it can predict the future movements of the other two indexes in a bull market. NASDAQ Composite is a technology-dominated index and there tend to be strong optimistic feelings in the stock market regarding possible future technological developments in a bull market. This can explain the leading role of the NASDAQ Composite index in bull markets. The test statistics in Table 4 indicate that the DJIA and S&P 500 indexes have the leading role and they can predict the future movements of the NASDAQ Composite Index in a bear market. In bear markets, the emphasis tends to be on economic fundamentals and the DJIA and S&P 500 indexes, dominated by the nation's largest companies, determine the direction of the market movements as the economy slides downward. The findings in this study imply that investors can earn above-normal profits by following the signals from the leading stock market indexes in bull and bear markets.

CONCLUDING COMMENTS

In this paper, we have studied the linkages between the DJIA, S&P 500, and NASDAQ Composite daily index returns during the July 17, 1998-March 24, 2000 and March 17, 2003-January 26, 2004 bull markets and the March 24, 2000-October 9, 2002 bear market.

The average daily returns of the three indexes are not statistically significant during the total five-year period covering both bull and bear markets. However, the NASDAQ Composite Index daily returns are significantly higher in a strong bull market and significantly lower in a bear market compared with the daily index returns of the other two indexes. The NASDAQ Composite is the most volatile index, and the DJIA is the least volatile index in both bull and bear markets. The NASDAQ Composite Index provides the highest average daily return per unit of volatility risk, and the DJIA Index provides the lowest average daily return per unit of volatility risk.

The rolling correlation analysis results indicate that the DJIA, S&P 500, and NASDAQ Composite daily index returns are generally highly positively correlated. However, the correlation between the daily returns of the indexes has considerable volatility over time. The correlation between the daily returns of the NASDAQ Composite Index and the other two indexes, particularly the S&P 500 Index, has a high negative value toward the end of the July 17, 1998-March 24, 2000 bull-market period, just before the beginning of the March 24, 2000-October 9, 2002 bear-market period. This may reveal an important information regarding the long-run co-movement behavior of the three major stock market indexes. A sharp sudden decrease in the correlation between the daily returns of the NASDAQ Composite Index and the other two indexes from a high positive value to a high negative value may signal the end of a strong bull market and the beginning of a severe bear market.

The Granger-causality test results indicate that the past daily returns of each of the three stock market indexes covered in this study can predict its own future daily returns both in a bull market and in a bear market, i.e., none of the three daily index returns follow a random walk. The past daily returns of the NASDAQ Composite Index can predict the future daily returns of the other two indexes in a bull market. In a bear market, however, the daily returns of the DJIA and S&P 500 indexes can predict the future daily returns of the NASDAQ Composite Index. Investors tend to be optimistic about future technological developments in a bull market. The NASDAQ Composite Index is dominated by high tech companies, and it appears to lead the other two indexes in a bull market. In a bear market, economic fundamentals play an important role and the DJIA and S$P 500 indexes, dominated by the nation's largest companies, lead the NASDAQ index as the economy slides downward. Investors can earn above-normal profits by following the signals from the leading indexes in bull and bear markets.

REFERENCES

Aggarwal, R., & Kyaw, N. A. (2005). Equity market integration in the NAFTA region: Evidence from unit root and co-integration tests. *International Review of Financial Analysis*, *14*(4), 393-406.

Enders, W. (1995). *Applied econometric time series.* New York: John Wiley and Sons.

Hilliard, J. E. (1979). The relationship between equity indices on world exchanges. *Journal of Finance*, *34*(1), 103-114.

Meric, I., Coopersmith, L. W., Wise, D., & Meric, G. (2002). The linkages between the world's major stock markets in the 2000-2001 bear market. *Journal of Investing*, *11*(4), 55-62.

Meric, I., & Meric, G. (1989). Potential gains from international portfolio diversification and inter-temporal stability and seasonality in international stock market relationships. *Journal of Banking and Finance*, *13*(4/5), 627-640.

Philippatos, G. C., Christofi, A., & Christofi, P. (1983). The inter-temporal stability of international stock market relationships: Another view. *Financial Management*, *12*(4), 63-69.

Ratner, M., & Leal, R. (1996). Causality tests for the emerging markets of Latin America. *Journal of Emerging Markets*, *1*(1), 29-40.

Solnik, B. H., Boucrelle, C., & Le Fur, Y. (1996). International market correlation and volatility. *Financial Analysts Journal*, *52*(5), 17-34.

Ilhan Meric, Lehigh University, is a professor of finance at Rider University teaching financial management. His research interests include global equity markets, portfolio analysis and corporate management.

Mitchell Ratner, Drexel University, is an associate professor of finance at Rider University teaching international finance and financial markets. His research interests include international finance and investments.

Gulser Meric, Lehigh University, is a professor of finance at Rowan University teaching financial management. Her research interests include global equity markets, investments and corporate finance.

Table 1. Average Daily Returns and the Volatility of Returns*

Panel A: Average Daily Returns and Their Volatility

	Bull Market[1]	Bear Market[2]	Bull Market[3]	Total Period[4]
DJIA				
Average Daily Return	0.000236	-0.000370	0.000963	0.000633
Standard Deviation	0.012575	0.013792	0.010249	0.012823
S&P 500				
Average Daily Return	0.000456	-0.000790	0.001042	0.000831
Standard Deviation	0.013404	0.014380	0.010374	0.013425
NASDAQ				
Average Daily Return	0.002266	-0.002190	0.001646	0.001570
Standard Deviation	0.019497	0.027475	0.013272	0.023103

Panel B: Comparison of Average Daily Returns[5]

	Bull Market[1,6]	Bear Market[2,7]	Bull Market[3,8]	Total Period[4,9]
DJIA vs. S&P 500				
t statistic	0.326	1.559	0.632	0.911
Significance	0.744	0.119	0.528	0.363
DJIA vs. NASDAQ				
t statistic	2.506	1.914	1.298	0.284
Significance	0.013	0.056	0.195	0.776
NASDAQ vs. S&P 500				
t statistic	3.173	1.866	1.360	0.034
Significance	0.002	0.063	0.175	0.973

* Daily return and standard deviation figures are measured in decimals.
[1] From July 17, 1998 to March 24, 2000
[2] From March 24, 2000 to October 9, 2002
[3] From March 17, 2003 to January 26, 2004
[4] All three periods combined.
[5] Two-tailed t-tests of matched samples.
[6] Degrees of freedom for the t-tests: 396.
[7] Degrees of freedom for the t-tests: 636.
[8] Degrees of freedom for the t-tests: 239.
[9] Degrees of freedom for the t-tests: 1,274.

Table 2. Average Rolling Correlation Coefficients Among the Daily Returns of the Indexes and the Volatility of the Correlation Coefficients*

	Bull Market[1]	Bear Market[2]	Bull Market[3]	Total Period[4]
DJIA and NASDAQ				
Correlation Coefficient**	0.51	0.68	0.82	0.65
Standard Deviation	0.48	0.30	0.21	0.37
Coefficient of Variation	0.95	0.44	0.26	0.58
DJIA and S&P 500				
Correlation Coefficient**	0.87	0.87	0.94	0.88
Standard Deviation	0.12	0.15	0.12	0.14
Coefficient of Variation	0.13	0.18	0.12	0.16
S&P 500 and NASDAQ				
Correlation Coefficient**	0.74	0.90	0.90	0.84
Standard Deviation	0.36	0.11	0.13	0.24
Coefficient of Variation	0.49	0.12	0.14	0.28

* Monthly rolling correlation between the daily returns of the indexes.
** The t-tests indicate that all correlation coefficients are significant at the one-percent level of significance.
[1] From July 17, 1998 to March 24, 2000
[2] From March 24, 2000 to October 9, 2002
[3] From March 17, 2003 to January 26, 2004
[4] All three periods combined.

Table 3. Granger-Causality Tests: Bull Market

Index	F-Statistic[1]	P-Value
	PANEL A: Dependent Variable: DJIA	
DJIA	821.9154*	0.0000
NASDAQ	6.4983*	0.0000
S&P500	2.1258	0.0607
	PANEL B: Dependent Variable: S&P500	
DJIA	1.1123	0.3525
NASDAQ	6.4434*	0.0000
S&P500	398.0870*	0.0000
	PANEL C: Dependent Variable: NASDAQ	
DJIA	0.9112	0.4732
NASDAQ	3789.9446*	0.0000
S&P500	1.2715	0.2745

* Significant at the one-percent level of significance.

Table 4. Granger-Causality Tests: Bear Market

Index	F-Statistic	P-Value
	PANEL A: Dependent Variable: DJIA	
DJIA	351.2324*	0.0000
NASDAQ	2.1000	0.0638
S&P500	2.1118	0.0624
	PANEL B: Dependent Variable: S&P500	
DJIA	2.0671	0.0678
NASDAQ	1.2739	0.2734
S&P500	312.9986*	0.0000
	PANEL C: Dependent Variable: NASDAQ	
DJIA	3.1658*	0.0079
NASDAQ	718.7919*	0.0000
S&P500	2.4783**	0.0309

* Significant at the one-percent level of significance.
** Significant at the five-percent level of significance.

Figure 1. Daily Returns of the DJIA, S&P 500, and NASDAQ Composite Indexes

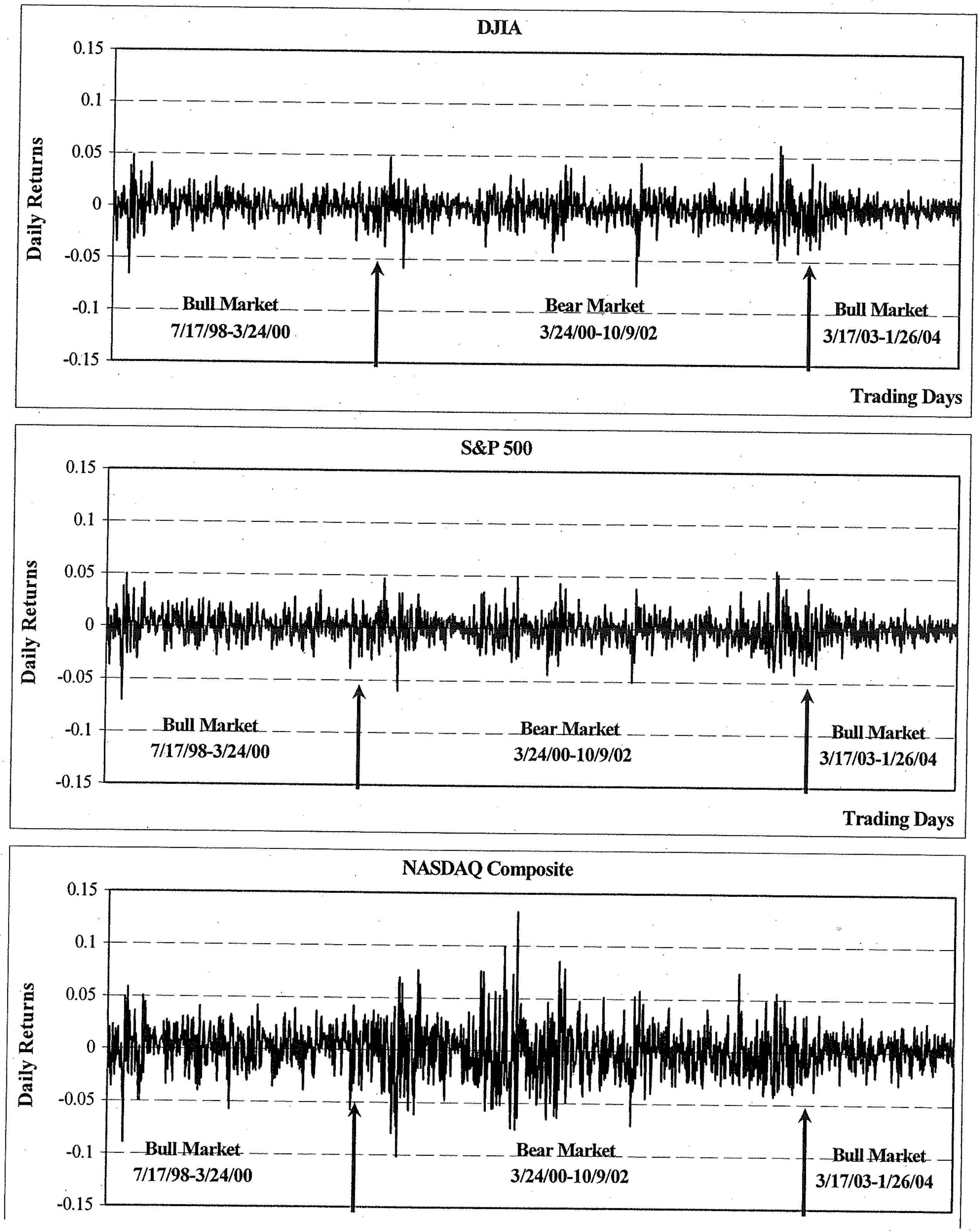
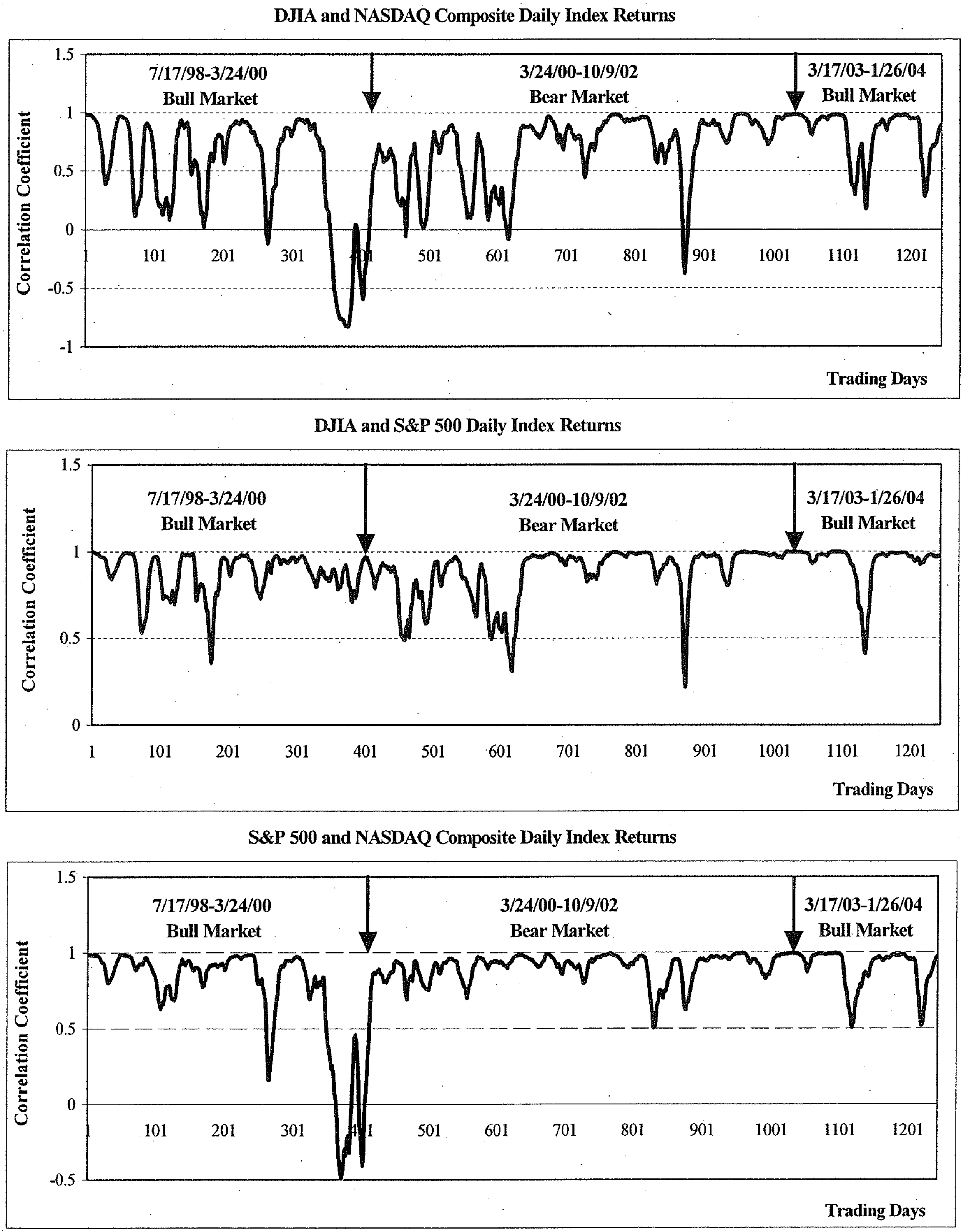


Figure 2. Monthly Rolling Correlation Among the Three Indexes



PENNSYLVANIA JOURNAL OF
BUSINESS AND ECONOMICS
Clarion University of Pennsylvania
College of Business Administration
Clarion, PA 16214