Philips, Thomas K. 2002. "The Source of Value." *Journal of Portfolio Management* 28 (4): 36–44.

Philips, Thomas K., and Ádám Kóbor. 2020. "Ultra-Simple Shiller's CAPE: How One Year's Data Can Predict Equity Market Returns Better Than Ten." *Journal of Portfolio Management* 46 (4): 140–55.

Philips, Thomas K., and Cenk Ural. "Uncloaking Campbell and Shiller's CAPE: A Comprehensive Guide to Its Construction and Use." *Journal of Portfolio Management* 43 (1): 109–25.

Phillips, Peter C. B., Shuping Shi, and Jun Yu. 2015. "Testing for Multiple Bubbles: Historical Episodes of Exuberance and Collapse in the S&P 500." *International Economic Review* 56 (4): 1043–78.

Reinganum, Marc. 1981. "Misspecification of Capital Asset Pricing: Empirical Anomalies Based on Earnings Yields and Market Values." *Journal of Financial Economics* 9 (1): 19–46.

Ross, Stephen A. 1987. "Regression to the Max." Unpublished manuscript, Yale University. www.researchgate.net/profile/Stephen_Ross3/publication/247378278_Regression_to_the_max/.

Samuelson, Paul A. 1958. "An Exact Consumption-Loan Model of Interest with or without the Social Contrivance of Money." *Journal of Political Economy* 66 (6): 467–82.

Shaw, Edward S. 1973. *Financial Deepening in Economic Development*. New York: Oxford University Press.

Sicard, Germain. 2015. *The Origins of Corporations: The Mills of Toulouse in the Middle Ages*. New Haven, CT: Yale University Press.

Siegel, Jeremy J. 1995. "The Nifty-Fifty Revisited: Do Growth Stocks Ultimately Justify Their Price?" *Journal of Portfolio Management* 21 (4): 8–20.

Siegel, Jeremy J. 2005. "Perspectives on the Equity Risk Premium." *Financial Analysts Journal* 61 (6): 61–71.

Siegel, Jeremy J. 2014. *Stocks for the Long Run*, 5th ed. New York: McGraw Hill.

Siegel, Jeremy J. 2016. "The Shiller CAPE Ratio: A New Look." *Financial Analysts Journal* 72 (3): 41–50.

Siegel, Jeremy J., and Jeremy D. Schwartz. 2006. "Long-Term Returns on the Original S&P 500 Firms." *Financial Analysts Journal* 62 (1): 18–31.

Straehl, Philip U., and Roger G. Ibbotson. 2017. "The Long-Run Drivers of Stock Returns: Total Payouts and the Real Economy." *Financial Analysts Journal* 73 (3): 32–52.

Straehl, Philip U., and Roger G. Ibbotson. 2018. "'The Long-Run Drivers of Stock Returns: Total Payouts and the Real Economy': Author Response." *Financial Analysts Journal* 74 (1).

van Binsbergen, Jules H., and Ralph S. Koijen. 2017. "The Term Structure of Returns: Facts and Theory." *Journal of Financial Economics* 124 (1): 1–21.

Named Endowments

CFA Institute Research Foundation acknowledges with sincere gratitude the generous contributions of the Named Endowment participants listed below.

Gifts of at least US$100,000 qualify donors for membership in the Named Endowment category, which recognizes in perpetuity the commitment toward unbiased, practitioner-oriented, relevant research that these firms and individuals have expressed through their generous support of CFA Institute Research Foundation.

Ameritech
Anonymous
Robert D. Arnott
Theodore R. Aronson, CFA
Asahi Mutual Life Insurance Company
Batterymarch Financial Management
Boston Company
Boston Partners Asset Management, L.P.
Gary P. Brinson, CFA
Brinson Partners, Inc.
Capital Group International, Inc.
Concord Capital Management
Dai-Ichi Life Insurance Company
Daiwa Securities
Mr. and Mrs. Jeffrey Diermeier
Gifford Fong Associates
John A. Gunn, CFA
Investment Counsel Association of America, Inc.
Jacobs Levy Equity Management
Jon L. Hagler Foundation
Long-Term Credit Bank of Japan, Ltd.
Lynch, Jones & Ryan, LLC
Meiji Mutual Life Insurance Company
Miller Anderson & Sherrerd, LLP
John B. Neff, CFA
Nikko Securities Co., Ltd.
Nippon Life Insurance Company of Japan
Nomura Securities Co., Ltd.
Payden & Rygel
Provident National Bank
Frank K. Reilly, CFA
Salomon Brothers
Sassoon Holdings Pte. Ltd.
Scudder Stevens & Clark
Security Analysts Association of Japan
Shaw Data Securities, Inc.
Sit Investment Associates, Inc.
Standish, Ayer & Wood, Inc.
State Farm Insurance Company
Sumitomo Life America, Inc.
T. Rowe Price Associates, Inc.
Templeton Investment Counsel Inc.
Frank Trainer, CFA
Travelers Insurance Co.
USF&G Companies
Yamaichi Securities Co., Ltd.

Senior Research Fellows

Financial Services Analyst Association

For more on upcoming CFA Institute Research Foundation publications and webcasts, please visit www.cfainstitute.org/research/foundation.

CFA Institute
Research
Foundation

Available online at www.cfainstitute.org

ISBN 978-1-952927-35-5

9 781952 927355 >

The Walk-down to Beatable Analyst Forecasts: The Role of Equity Issuance and Insider Trading Incentives*

SCOTT RICHARDSON, *University of Pennsylvania*

SIEW HONG TEOH, *The Ohio State University*

PETER D. WYSOCKI, *Massachusetts Institute of Techonology*

Abstract

It has been alleged that firms and analysts engage in an "earnings-guidance game" where analysts first issue optimistic earnings forecasts and then "walk down" their estimates to a level that firms can beat at the official earnings announcement. We examine whether the walk-down to beatable targets is associated with managerial incentives to sell stock after earnings announcements on the firm's behalf (through new equity issuance) or from their personal accounts (through option exercises and stock sales). Consistent with these hypotheses, we find that the walk-down to beatable targets is most pronounced when firms or insiders are net sellers of stock after an earnings announcement. These findings provide new insights on the impact of capital-market incentives on communications between managers and analysts.

Keywords Analysts' forecasts; Earnings guidance; Insider trading; New equity issuance; Stock options

JEL Descriptors G14, G30, G38, K22, M41

* Accepted by Greg Waymire. We thank an anonymous referee, conference discussants Afshad Irani, Laura Starks, and Kent Womack, and Lisa Bryant, Patricia Dechow, Peter Easton, David Hirshleifer, Doug Skinner, Stephen Taylor, Irem Tuna, Greg Waymire, Eric Wruck, Yinglei Zhang, and conference participants at the American Accounting Association meetings, American Finance Association meetings, Berkeley area research talks, and the 1999 Texas Finance Festival, and seminar participants at the University of Kansas, London School of Economics, Mellon Bank Capital Management, and Ohio State University. We thank I/B/E/S (a service of Thomson Financial) for providing analyst forecast data as part of a broad academic program to encourage earnings-expectation research. This paper was previously circulated under the title "Trading Analysts' Forecasts over the Annual Earnings Horizon: Are Analysts' Forecasts Optimistic or Pessimistic?"

La réévaluation des prévisions des analystes à des niveaux permettant le dépassement : le rôle de l'émission d'actions et des facteurs incitatifs aux délits d'initiés

Condensé

Certains prétendent que les sociétés et les analystes se livrent à un « exercice de guidage des résultats » dans lequel les analystes produisent d'abord des prévisions de résultats optimistes pour revenir ensuite sur leurs estimations et les ramener à un niveau que les sociétés sont en mesure de dépasser lors de l'annonce officielle de leurs résultats. Les auteurs élaborent et testent des hypothèses relatives à ce passage des analystes de l'optimisme au pessimisme, à partir des facteurs qui incitent les dirigeants à vendre les actions de la société à des conditions avantageuses en évitant de décevoir les investisseurs lors de l'annonce officielle des résultats de l'entreprise.

L'analyse des auteurs repose sur cinq éléments sous-jacents à l'exercice de guidage des résultats. Premièrement, dans la majorité des opérations, les ventes d'actions par les dirigeants et par l'entreprise se déroulent sur un court laps de temps après les annonces de résultats. Deuxièmement, les dirigeants qui ont l'intention de vendre des actions pour leur propre compte ou au nom de la société après une annonce de résultats s'intéressent au cours des titres de la société à brève échéance après l'annonce. Troisièmement, les dirigeants peuvent influencer les analystes dans leurs prévisions de résultats grâce à la publication d'informations discrétionnaires, et les analystes sont, pour leur part, enclins à collaborer. Quatrièmement, les analystes tendent généralement à être optimistes dans leurs prévisions initiales. Enfin, le marché paraît gratifier les sociétés qui dépassent les dernières prévisions de résultats des analystes d'évaluations supérieures à celles qu'il octroie aux entreprises qui ne sont pas parvenues à dépasser l'objectif prévisionnel, peu importe la voie ou le moyen emprunté pour atteindre l'objectif (soit le guidage des anticipations ou la gestion des résultats). À partir de ces éléments, les auteurs font l'hypothèse que les dirigeants guident systématiquement les analystes vers des objectifs prévisionnels qui peuvent être dépassés, de sorte qu'eux-mêmes ou leurs sociétés puissent vendre des actions à des conditions avantageuses après une annonce de résultats.

Les auteurs exposent d'abord des faits qui relient l'évolution du profil des prévisions des analystes entre les années 1980 et les années 1990 et les changements institutionnels et réglementaires qui ont accentué les facteurs liés au marché financier incitant les dirigeants à guider les analystes dans leurs prévisions de résultats et à dépasser ces objectifs prévisionnels, afin de hausser le cours des actions. Ces changements systémiques incluent l'utilisation accrue de la rémunération des dirigeants sous forme d'options sur actions, la restriction des négociations par les initiés à la période postérieure aux annonces de résultats en réponse à l'*Insiders' Fraud and Securities Trading Act* de 1988 et le remaniement, en 1991, de la règle relative au délai d'attente que doivent respecter les initiés entre les opérations de négociation (« *short-swing rule* »), de façon à leur permettre de lever leurs options et de vendre immédiatement les actions de la société. L'analyse des auteurs montre qu'entre 1984 et 2001, les prévisions de résultats initiales trimestrielles et annuelles des analystes sont trop optimistes par rapport aux résultats réels finals. Lorsque la date de l'annonce des résultats approche, les analystes révisent à la baisse leurs prévisions afin qu'elles soient moins optimistes par rapport aux résultats réels. Il existe une différence essentielle entre les années

1980 et les années 1990 : les révisions moyennes et médianes des prévisions de résultats des analystes au cours de la période s'échelonnant du milieu jusqu'à la fin des années 1990 deviennent bel et bien pessimistes lorsque la date de l'annonce des résultats approche. Ce virage systématique des analystes vers le pessimisme dans les années 1990 coïncide avec les changements institutionnels et réglementaires qui ont accentué les facteurs liés au marché financier incitant les dirigeants à guider les analystes dans leurs prévisions de résultats et à dépasser ces objectifs prévisionnels, afin de hausser le cours des actions à brève échéance.

Les auteurs soumettent à des tests transversaux leur prédiction principale selon laquelle les facteurs incitatifs liés au marché financier découlant de la vente d'actions, soit à titre personnel (la levée d'options et la vente d'actions par les initiés) soit au nom de la société (l'émission de nouvelles actions), sont associés au fait que les analystes ramènent leurs prévisions à un niveau que les sociétés sont en mesure de dépasser. Dans leurs tests transversaux, les auteurs utilisent un vaste échantillon de prévisions des analystes, du milieu des années 1980 jusqu'à 2001, tirées de la base de données I/B/E/S. Les données sur la vente d'actions par les dirigeants sont tirées de la compilation, effectuée par la société Thompson Financial, des opérations d'initiés soumises à la SEC. Seules les opérations des initiés parmi les achats et les ventes sur le marché libre et la levée d'options figurent dans le calcul des ventes nettes d'actions par les dirigeants. Les auteurs mesurent les ventes d'actions au nom de la société en utilisant les données relatives aux émissions d'actions dans le trimestre au cours duquel sont annoncés les résultats et le trimestre subséquent.

Conformément à leur principale prédiction transversale, les auteurs constatent que le pessimisme dans les prévisions antérieures à l'annonce de résultats est le plus marqué dans le cas des sociétés dont les dirigeants sont le plus fortement incités par les facteurs liés au marché financier à éviter les déceptions relatives aux résultats. Les auteurs observent que les sociétés dont les dirigeants vendent des actions après une annonce de résultats sont plus susceptibles d'être associées à des prévisions pessimistes des analystes avant l'annonce des résultats. La probabilité de pessimisme des prévisions passe de 54 %, dans le cas d'une société moyenne pour laquelle n'est enregistrée aucune vente nette par les initiés, à 66 % dans le cas d'une société moyenne pour laquelle est enregistrée une vente nette subséquente par les initiés. En outre, les sociétés dont les initiés sont des vendeurs nets d'actions de l'entreprise sont également plus susceptibles d'être associées à des analystes qui passent de l'optimisme à long terme au pessimisme à court terme avant l'annonce de résultats. La probabilité du passage de l'optimisme, tôt dans le trimestre, au pessimisme, à proximité de l'annonce des résultats, augmente de 21 % chez les sociétés pour lesquelles n'est pas enregistrée de vente nette des initiés à 27 % chez les sociétés pour lesquelles est enregistrée une vente nette des initiés. Cette constatation est conforme au fait que les dirigeants orientent les analystes vers des prévisions de résultats pouvant être dépassées pour faciliter les opérations avantageuses que peuvent conclure les initiés après les annonces de résultats.

Les auteurs constatent que les résultats de leur série chronologique résistent : 1) à différents déflateurs des prévisions de résultats des analystes, 2) aux horizons prévisionnels annuel aussi bien que trimestriel, 3) à l'utilisation de la population entière des sociétés figurant dans la base de données I/B/E/S et à l'utilisation d'un échantillon déterminé de sociétés examinées durant toute la période étudiée et 4) aux ajustements visant la prise en compte des fractionnements d'actions susceptibles d'influer sur le calcul des erreurs prévisionnelles des analystes.

Ils constatent également que leurs résultats empiriques transversaux résistent : 1) à différents déflateurs des prévisions de résultats des analystes, 2) aux horizons prévisionnels annuel aussi bien que et trimestriel, 3) à l'inclusion de diverses caractéristiques des sociétés précédemment liées aux prévisions de résultats des analystes, 4) aux différents types d'analystes (précurseurs ou retardataires) et 5) aux différentes classes d'investisseurs, y inclus les investisseurs institutionnels et les investisseurs individuels.

Les constatations des auteurs complètent les résultats d'Aboody et Kasznik (2000) dont les observations confirment que les dirigeants publient de l'information à des fins stratégiques, en vue d'obtenir des options sur actions à des conditions avantageuses. L'approche des auteurs consiste à examiner les facteurs qui incitent les dirigeants à publier de l'information à des fins stratégiques dans le but de lever des options et de vendre des actions à des conditions avantageuses. Ils poussent également plus loin les études récentes portant sur les caractéristiques des sociétés qui se livrent au guidage des résultats (Matsumoto, 2002) en analysant explicitement les facteurs qui incitent directement les dirigeants à tirer profit de ce guidage. Pour conclure, les résultats empiriques de l'étude nous renseignent davantage sur l'incidence des facteurs incitatifs liés au marché financier sur les communications entre dirigeants et analystes.

1. Introduction

Security regulators and the business press have often alleged that firms and analysts are involved in an "earnings-guidance game". These critics claim that analysts issue systematically optimistic earnings forecasts at the start of the fiscal period and then "walk down" their estimates to a level the firm can beat on the formal earnings announcement. For example, Laderman (1998, 148) noted in a *Business Week* article:

> Thanks to the IR [investor relations] people and analysts, in recent years, earnings estimates for the S&P 500 in any quarter tend to start out an average 5% to 8% higher than where the earnings end up. The Street knows this and allows for analysts to whittle down the numbers as the quarter proceeds.

We develop and test hypotheses about this pattern of analyst optimism-to-pessimism based on managerial incentives to sell company stock on favorable terms by avoiding a "disappointment" on the official announcement of firm earnings. The motivation for our investigation is straightforward. As Ken Brown (2002, C1) indicates in his *Wall Street Journal* column:

> the reasons that executives became so obsessed with hitting their numbers are clear. A company that shows steady growth with few surprises often gets rewarded with a sweet premium from investors — a high stock price — which goes a long way toward keeping the executives' stock options in the money.

The business press is replete with articles alleging that firms deliberately attempt to deceive or pressure analysts into issuing "beatable" earnings targets. Even as far back as May 6, 1991, Laurie P. Cohen, staff reporter of the *Wall Street Journal* wrote that

> after securities analysts estimate what the companies they follow will earn, the game begins. Chief financial officers or investor-relations representatives traditionally give "guidance" to analysts, hinting whether the analysts should raise or lower their earnings projections so the analysts won't be embarrassed later.
>
> And these days, many companies are encouraging analysts to deflate earnings projections to artificially low levels, analysts and money managers say. If the game is played right, a company's stock will rise sharply on the day it announces its earnings — and beats the analysts' too conservative estimates.

Prior academic research documents that analysts issued systematically optimistic forecasts during the 1980s (see, e.g., O'Brien 1988). However, consistent with media reports of forecast pessimism, more recent empirical evidence suggests that firms attempt to meet or beat earnings-forecast benchmarks (see, e.g., Bartov, Givoly, and Hayn 2002; Burgstahler and Eames 2002; DeGeorge, Patel, and Zeckhauser 1999; Kasznik and McNichols 2002; Matsumoto 2002; and Richardson, Teoh, and Wysocki 1999). In this paper, we explore empirically whether capital-market incentives stemming from the sale of equity either on personal account (insider option exercise and stock sale) or on the firm's behalf (new equity issuance) are associated with the walk-down of analysts' forecasts to targets that are eventually beaten through successful guidance of expectations or earnings management.

We begin our analysis by developing a framework for the earnings-guidance game. The framework is based on five underlying elements outlined below, and discussed in more depth in section 2. First, in the majority of transactions, managerial and firm equity sales occur during a short window after earnings announcements. Second, managers who are about to sell shares on their personal account or on behalf of the firm after an earnings-announcement care about the firm's short-term post-announcement stock price level. Third, managers can influence analysts' earnings targets through discretionary information disclosures and analysts have incentives to cooperate. Fourth, analysts' initial forecasts generally tend to be optimistic. Finally, the market appears to reward firms that beat analysts' latest earnings target with higher valuations than those that fail to beat the target, regardless of the path to the target or how the target is achieved (that is, through guiding expectations or earnings management). On the basis of these elements, we hypothesize that managers systematically guide analysts toward beatable targets so that they or their firms can sell equity on favorable terms after an earnings announcement. According to this managerial guidance hypothesis, such guidance allows the manager to maintain favorable stock market valuations exactly when they are needed, just after earnings announcements.

In our empirical study, we test this hypothesis by examining the association between firms' and managers' equity sales after earnings announcements and (1) the walk-down in analysts' optimistic forecasts early in the fiscal period and (2) firms meeting or beating analysts' final revised earnings targets. Given that neither managers' intentions to guide analysts nor their communications with analysts can be directly observed in our sample, we follow prior empirical studies of agency models and examine principals' and agents' observable actions, after controlling for other influences.[1] In our study, the analysts' observable actions are their beatable forecast revisions and the managers' observable actions are their post-earnings announcement equity transactions. Our evidence is consistent with the predictions of our managerial guidance hypothesis, whereas alternative interpretations do not appear to explain the totality of our results.[2]

In our tests, we use a large sample of analyst forecasts from the mid-1980s to 2001 available from I/B/E/S. Data on managers' sale of shares are obtained from Thomson Financial's compilation of insider trades that are filed with the Securities and Exchange Commission (SEC). Only insiders' trades from open-market purchases and sales and option exercises are included in the calculation of the net sale of shares by the managers. We measure the sale of shares on the firm's own behalf using data on equity issuances in the quarter of and quarter after the earnings announcement.

Consistent with our main predictions, we find that analysts' earnings forecast pessimism prior to an earnings announcement is (1) more prevalent in the late 1990s following institutional and regulatory changes that increased managers' capital-market incentives to guide and beat analysts' forecasts to boost short-term stock prices, and (2) more common for firms that are about to issue new equity and whose insiders are net sellers of the firm's stock in the quarter immediately following an earnings announcement.

Our findings complement the results of Aboody and Kasznik 2000, who present evidence consistent with managers' strategically disclosing information in order to obtain stock options on favorable terms. Our approach examines managerial incentives to strategically disclose information in order to exercise options and sell stock on favorable terms. We also contribute to the recent literature (e.g., Matsumoto 2002) examining firm characteristics that influence earnings guidance by explicitly considering firm and managers' direct incentives to profit from earnings guidance in our study.

The rest of the paper is structured as follows. In section 2, we develop our hypotheses. Section 3 describes the sample and data. Section 4 presents descriptive evidence for the behavior of earnings forecasts over the fiscal period in various calendar subperiods. In section 5, we present primary cross-sectional tests and a robustness analysis of the predictions arising from the earnings-expectations game. Section 6 concludes the paper.

2. Background and hypothesis development

In this section, we motivate the prediction that managers' capital-market trading incentives are related to their guidance of analysts' earnings forecasts. We first discuss the institutional rules governing the timing of stock-sale transactions that motivate managers to focus on the firm's stock price around earnings announcements. We then discuss how analysts' forecasts influence stock prices, suggest why analysts cooperate with managers in setting forecasts, and discuss recent empirical research consistent with managers' influencing analysts' forecasts. Finally, we discuss recent research indicating that investors fixate on meeting earnings thresholds such as analysts' forecasts and reward good versus bad news asymmetrically. We argue that if the market rewards firms that beat analysts' latest earnings target and if managers wish to sell equity on favorable terms after earnings announcements, then managers have strong incentives to influence analysts' expectations to avoid an earnings disappointment. We combine these elements to develop hypotheses on the cross-sectional variation in analysts' optimism and pessimism. Together, these elements suggest that insider trading and new equity issuance activities are linked to analyst forecast bias within the fiscal period.

Why and when managers care about short-term stock price

Managers intending to issue new equity on the firm's behalf care about the firm's stock price level after an earnings announcement because the stock price directly affects the proceeds the firm can raise through an equity sale. Managers care particularly about the stock price right after an earnings announcement because new equity issues typically occur in the weeks following a public earnings announcement (see, e.g., Korajczyk, Lucas, and MacDonald 1991). Lucas and MacDonald (1990) explain this timing as an attempt to minimize information asymmetry between the firm and uninformed outside investors by delaying equity issues until after an earnings announcement.

Stock-based compensation such as stock options also motivates managers to care about the firm's stock price by directly tying compensation to the firm's stock price performance.[3] Hall and Liebman (1998) report that stock options have become an increasingly important portion of managers' compensation. They report that stock option grants increased to make up almost 50 percent of chief executive officer (CEO) compensation by 1994. Thus, managers face increasing incentives to care about the firm's stock price from the structure of their compensation package.

Furthermore, managers care about the firm's short-term stock price specifically during the earnings-announcement period because of institutional constraints on insider trading. These restrictions have arisen because regulatory and corporate concerns that managers may use their inside information to exercise stock options or trade in the firms' stock at the expense of outside investors. U.S. insider trading laws (Insider Trading Sanctions Act 1984; Insider Trading and Securities Fraud Enforcement Act 1988) expressly prohibit this direct profit-taking opportunity by insiders. In response to the 1988 Insider Trading and Securities Fraud Enforcement Act, firms increasingly have instituted their own policies and procedures to regulate trading by insiders prior to earnings announcements. These restrictions generally take the form of explicit blackout periods specifically in the last two months before the earnings-announcement date (see, e.g., Bettis, Coles, and Lemmon 2000; Jeng 1999). Bettis et al. reported that firms increasingly instituted formal blackout periods during the 1990s, and that by 1997, 80 percent of firms had blackout periods.[4] Therefore, the occurrence of insiders' option exercises and stock sales are increasingly focused in a narrow window immediately after an earnings announcement. Consistent with this, Sivakumar and Waymire (1994) report a higher incidence of insider trades in the week immediately after a quarterly earnings announcement. Similarly, Noe (1999) reports that insider transactions cluster after voluntary disclosures that are favorable to stock prices.

In sum, stock option compensation, insider trading restrictions, and new equity issue guidelines motivate managers to care about the firm's short-term stock price immediately following an earnings announcement. As a result, the stock price level during the earnings-announcement period carries special significance for firm management.

Managers' ability to manage analysts' forecasts and analysts' incentives to cooperate

Empirical and anecdotal evidence suggest that managers can indeed influence analysts' earnings forecasts. As a key provider of information to analysts, managers can affect analysts' earnings expectations by controlling the content and timing of discretionary information releases. Soffer, Thiagarajan, and Walther (2000) find that firms use pre-announcements of earnings to manage analysts' expectations. They also find that managers are selective in the content of their disclosures and appear to receive stock price benefit from managing analysts toward beatable targets. Cotter, Tuna, and Wysocki (2004) find that the switch to pessimistic forecasts appears to be concentrated around the release of management forecasts. Using survey data, Hutton (2003) finds that firms where managers indicated that they provide active guidance to analysts are less likely to experience negative earnings surprises. Together these papers suggest that managers are both able and willing to engage in expectations management.

Francis and Philbrick (1993) and Lim (2001) argue that managers can pressure analysts to revise forecasts away from their true beliefs because of analysts' dependence on management for future information. The business press has reported incidents of analysts who issued unfavorable forecasts being shunned by the management. Analysts may find it very difficult to do their jobs if they are ignored by management at investor conferences and if the firm does not return analysts' phone calls for information. At the extreme, there have been allegations of analysts losing their jobs after writing negative reports about favored clients.

It has also been alleged that analysts face conflicting incentives in maintaining the quality of investment research versus securing investment banking deals. Laderman (1998) asserts that

> [m]ost Wall Street research is pitched to institutional investors who pay the firm about a nickel a share in commissions. But if an analyst spends his time trying to land an initial public offering, the firm can earn 15 to 20 times that amount per share. Investment banking deals are much more lucrative for the brokerage firm. Merger advisory fees can be sweet as well … . But what happens when there's a conflict between objective analyses and the demands of investment bankers? … There's no conflict. That's been settled. The investment bankers won.

It is a widespread belief in the business press and among regulators that highly lucrative underwriting deals often pressure analysts to cooperate with firms issuing new securities. The SEC's investor education website specifically mentions the potential for analyst conflict of interest because of investment banking relationships. The recent well-publicized $1.4 billion settlement between 10 major brokerages and the U.S. securities regulators stems from this very allegation that investment banking influences compromise analysts' objectivity. The legal investigation revealed many instances where analysts yielded to investment banking business pressures. The new Regulation AC, released by the SEC in April 2003, specifically requires a research analyst to certify that "the views expressed in the research report accurately reflect such research analyst's personal views". It also requires analysts to certify that his or her compensation was not directly or indirectly related to the recommendation; if it was, the extent and source of the relation must be disclosed in the report.[5]

Previous academic research has also provided some evidence that analysts yielded to client firm pressures. Collectively, Lin and McNichols (1998), Michaely and Womack (1999), Dechow, Hutton, and Sloan (2000), Teoh and Wong (2002), and Bradshaw, Richardson, and Sloan (2003) provide evidence that analysts' recommendations, forecasts, and price targets are biased because of the conflict of interests introduced by external financing and the associated potential for underwriting business.

General optimism in long-horizon forecasts

To have a walk-down from optimism to pessimism as the forecast horizon shortens, there needs to be optimism at long horizons. All past empirical studies on earnings forecasts have found systematic analyst optimism at long horizons, and we confirm this for our sample in both earlier and more recent periods. Our hypothesis is potentially consistent with different possible reasons for the pervasive initial optimism.

One possibility is an agency problem wherein analysts, on behalf of firms, make high forecasts in order to improve market perceptions of the firms.[6] The analysts benefit from covering firms that subsequently do well, so there may be a self-selection tendency for analysts to cover firms about which they are optimistic (see McNichols and O'Brien 1997). Alternatively, analysts could simply be irrationally prone to optimism. Regardless of the source of the initial optimism, our hypothesis is based on the presence of a distinct force acting toward pessimism just before earnings announcements.

Managers' incentives to achieve beatable targets

In addition to long-horizon forecast optimism, past studies have shown increased forecast accuracy as the earnings-announcement approaches. However, this research has generally found continued analyst optimism at all forecast horizons (see, e.g., Brown, Foster, and Noreen 1985). As discussed in the introduction, it is only in more recent periods that researchers have found evidence of analyst pessimism in short horizons. These authors suggest that management communications with analysts lead to the deflated earnings expectations.

Systematic analyst optimism implies that firms are more likely to miss rather than beat analysts' targets. This can have detrimental effects for a firm if investors' perception of the firm is influenced by whether it meets certain earnings thresholds.

For example, Skinner and Sloan (2002) find an asymmetry in investor reaction to beating versus missing a threshold consisting of analyst forecasts made in the last month prior to the earnings announcement. They find that when firms fall short of forecasts, the stock price drops more than the stock price rises when firms beat forecasts by an equivalent magnitude of earnings surprise. They also find that this asymmetry is especially pronounced for high-growth firms. The discontinuity in investor reaction to missing versus meeting or beating analysts' forecasts creates incentives for managers to guide analysts to beatable earnings forecasts prior to an earnings announcement. A slightly lower forecast can cause the firm to barely beat the forecast instead of missing it, which significantly increases the firm's expected post-earnings-announcement stock price.

Kasznik and McNichols (2002) and Bartov, Givoly, and Hayn (2002) find that the capital market provides a valuation premium to firms whose earnings meet or beat analysts' estimates. Specifically, Bartov, Givoly, and Hayn (2002, 196) find that the capital-market premium for meeting or beating forecasts remains significant after controlling for the overall earnings performance in the quarter and even despite the earlier dampening of expectations by earnings guidance. Their further tests provide evidence that the market-valuation premium persists for firms that meet or beat analysts' earnings forecasts that were revised late in the quarter. In other words, the path by which analyst forecasts come to be beaten appears to be less crucial than whether the forecast ultimately becomes beatable just prior to the earnings announcement, consistent with investor limited attention about the shifting benchmark.

Institutional forces and incentives to beat targets

Two structural changes between the 1980s and 1990s are likely to have increased managerial incentives to guide analysts toward beatable earnings targets. The first structural change is the greater use of stock-based executive compensation by U.S. corporations during the 1990s. For example, Hall and Liebman (1998) present evidence on the growing use of CEO stock option compensation in the 1990s as compared with the 1980s. The mean salary and bonus in 1994 was \$1.3 million and the mean value of stock options was \$1.2 million. Between 1980 and 1994, mean salary and bonus grew 97 percent whereas mean stock option value grew by over 680 percent. Murphy (1999) confirms this growth and shows that the explosive growth trend in stock options continued to 1996, the latest year in his study. The greater predominance of exercisable stock options in the 1990s encouraged greater managerial attention to stock prices, especially around the earnings-announcement date, given the insider-trading restrictions mentioned earlier. This increase in managerial stock sales after earnings announcements in the 1990s likely led to widespread incentives for managers to guide analysts' earnings forecasts to avoid any disappointments that would negatively affect share prices.[7]

The second structural change occurred in May 1991, when securities regulators changed the "short-swing rule" affecting insiders' stock option exercises. Prior to 1991, section 16b of the Securities Exchange Act of 1934 required insiders to hold shares of stocks acquired through an option exercise for at least six months before selling, or the profits would go to the firm. In May 1991, the SEC effectively removed this restriction by changing the starting date of the six-month holding period from the exercise date to the option grant date. Consequently, since May 1991, managers have a more precise target date for when to exercise their stock options and immediately unload their stock, typically in the trading window after earnings announcements. Thus, the incentives to avoid an earnings disappointment by guiding forecasts to a beatable target increased subsequent to 1991.

Hypotheses on cross-sectional determinants of analyst pessimism

To summarize, the key elements that are related to the expectations-management game are that managers care about short-term share prices if they are about to sell shares on their personal account or on behalf of the firm after an earnings announcement, that managers can influence analysts' expectations through their information disclosures, and that the market appears to reward firms that beat analysts' latest earnings targets. Therefore, managerial incentives to guide analysts' forecasts are strongest if the firm and/or its managers are about to sell stock. This leads to the following cross-sectional prediction:

HYPOTHESIS 1. *The likelihood of observing short-horizon pessimistic analyst forecasts prior to an earnings announcement is increasing in management and firm incentives to sell stock after an earnings announcement. These effects are likely to be stronger in the 1990s than in earlier periods.*

Finding evidence in support of this hypothesis is consistent with analysts' being guided toward a more pessimistic target. However, another way to interpret the correlation between post-earnings-announcement equity sales and short-horizon pessimism is that stockholders sell shares after truly unexpected good news. If managers guide analysts toward beatable targets, then a stronger prediction can be derived on the basis of the following: (1) analysts initially issue optimistic (or unbiased) earnings forecasts, (2) analysts then revise their forecasts to become pessimistic before an earnings announcement, and (3) the firm or its insiders sell stock after the firm beats the revised earnings target. Therefore, we should observe an "opportunistic" switch from optimistic (or unbiased) to pessimistic analyst forecasts prior to firm or insider equity sales.[8] This leads to our second more restrictive prediction on cross-sectional determinants of expectations management:

HYPOTHESIS 2. *The likelihood of observing a switch from optimistic to pessimistic analyst forecasts prior to an earnings announcement is increasing in management and firm incentives to sell stock after an earnings announcement. These effects are stronger in the 1990s than in earlier periods.*

3. Sample and variable construction

Data on individual analysts' forecasts of quarterly and annual earnings per share are obtained from the Institutional Brokers Estimate System (I/B/E/S) Detail

History U.S. Edition tapes from 1984 to 2001. Unlike many previous studies, we use individual analysts' forecasts to calculate consensus forecasts to avoid potential staleness of the I/B/E/S consensus forecasts (see, e.g., Abarbanell and Bernard 1992).[9] The data sample consists of all individual analyst forecasts for firms with data availability on both I/B/E/S and COMPUSTAT.[10] To track forecast revisions leading up to the earnings' announcement, we sort analysts' forecasts into groups by 30-day blocks prior to the earnings release date over the annual horizon, and into finer two-week blocks over the quarterly horizon in the I/B/E/S Actuals File. We calculate a 30-day (or two-week) consensus forecast for each firm using the median of individual analyst forecasts within a period. We ensure that the calculation of the period's initial consensus forecast is made after the prior period's earnings announcement.

The forecast error (*FE*) is defined as the actual earnings per share minus the median forecast of earnings per share scaled by the stock price at the beginning of the quarter. The stock price deflator is used to control for potential spurious relations resulting from cross-sectional scale differences in earnings per share.[11] A negative error implies an optimistic forecast (that is, bad news), whereas a positive error implies a pessimistic forecast (that is, good news). Formally, the scaled forecast error (*FESC*) for firm i in quarter q and forecast-horizon period $-t$ is calculated as:

$$FESC_{i,q,t} = [Actual\ EPS_{i,q} - Forecast\ EPS_{i,q,t}]/P_{i,q-1} \qquad (1).$$

Firms' actual earnings per share are obtained from I/B/E/S for comparability with the forecast. The deflator $P_{i,q-1}$ is the stock price when the first forecast is available on I/B/E/S for firm i in quarter q. For annual forecasts, the deflator is the first available stock price in the year reported in I/B/E/S, which is typically available 12 months prior to the actual earnings-announcement date.[12] For quarterly forecasts, the deflator is the first available stock price in the quarter reported in I/B/E/S, which is typically available 3 months prior to the actual earnings-announcement date. To remove the influence of extreme outliers due to data-coding errors, we remove the extreme forecast errors that are greater than 10 percent in absolute value of share price.[13]

4. Pattern of forecast bias over the fiscal horizon

In section 2, we described how significant structural changes in executive compensation and insider-trading policies may affect managerial trading incentives in the 1990s, and consequently increased managerial incentives to guide analysts' forecasts. Before testing for a relation between managers' trading behavior and forecast revisions, we first examine temporal changes in analysts' forecast bias in the period from 1986 to 2001.

Panel A of Figure 1 shows the dynamic pattern of forecast bias over the annual forecast horizon for five calendar subperiods: 1984–88, 1989–91, 1992–94, 1995–97, and 1998–2001. For each subperiod, the forecasts show a consistent walk-down pattern. All subperiod initial median forecasts are optimistic, and the forecasts become increasingly less optimistic as the horizon shrinks toward the announcement date. A key difference across subperiods is that the median forecast crosses over to become pessimistic toward the earnings-announcement date only for the later calendar subperiods in the 1990s, consistent with the institutional changes noted for the 1990s. Furthermore, the median forecasts become pessimistic earlier in the forecast horizon as the 1990s progressed. For example, the median forecast becomes pessimistic in *Month −2* for the 1992–94 period, and in *Month −3* for 1995–97 and 1998–2001 subperiods. These findings are mirrored in the quarterly forecast data depicted in panel B of Figure 1. In this panel, one gets a more detailed picture of the short-horizon shift to pessimistic forecasts using two-week windows just prior to quarterly earnings announcements. Again, the shift to pessimism is only evident in the 1990s for the quarterly horizon.

The dynamic patterns of a shift toward pessimistic forecasts over the forecast horizon and over calendar subperiods are robust with respect to the empirical measures of forecast pessimism. For example, similar patterns are observed using mean analyst forecast errors. More important, our focus on the median forecasts indicates that the dynamic pattern of forecast bias documented here is independent of the debate on whether the mean forecast is biased.

The median forecast error in *Month 0* is only one cent in the post 1992 subperiods. The small magnitude does not imply low economic significance because "just beating" the forecast may have disproportionate informational signaling value to investors (see, e.g., DeGeorge et al. 1999). Overall, the univariate results present compelling evidence of a switch to systematic pessimism that is coincident with increased use of executive stock option compensation, greater concentration of insider trades in the post-earnings-announcement period, and the lifting of the short-swing rule for insiders during the 1990s.

Robustness checks on the temporal pattern

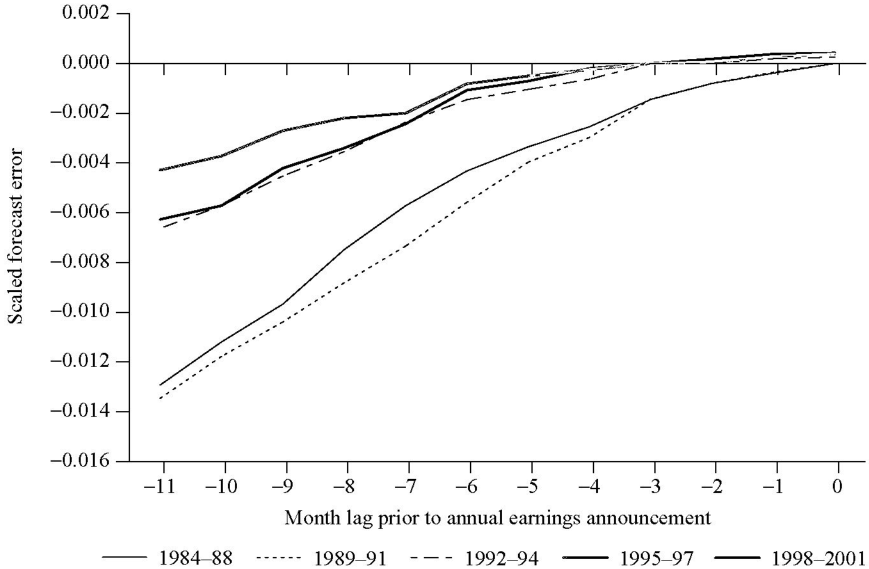
The analyst forecast errors in our sample are price-deflated to allow direct comparison across firms, which is standard in the literature. Given that scaling by price may introduce intertemporal variation in forecast bias if price–earnings ratios change over time, we also perform the tests using total assets per share as an alternative deflator. Our findings are robust using this alternative deflator. Figure 1 documents a switch in forecast error from optimism to pessimism as the horizon moves toward the earnings announcement in the subperiods after 1991. Note that the sign switch from optimism to pessimism forecasts is independent of the deflator because both price and total asset deflators are positive.

We also considered whether the time-series patterns are affected by changing sample composition during the sample period. For example, a change in the composition of publicly traded companies or in the breadth of coverage on I/B/E/S may affect the forecast bias over time. To rule this out, we replicated our tests using a constant sample of firms that existed throughout the sample period and found a similar dynamic pattern.

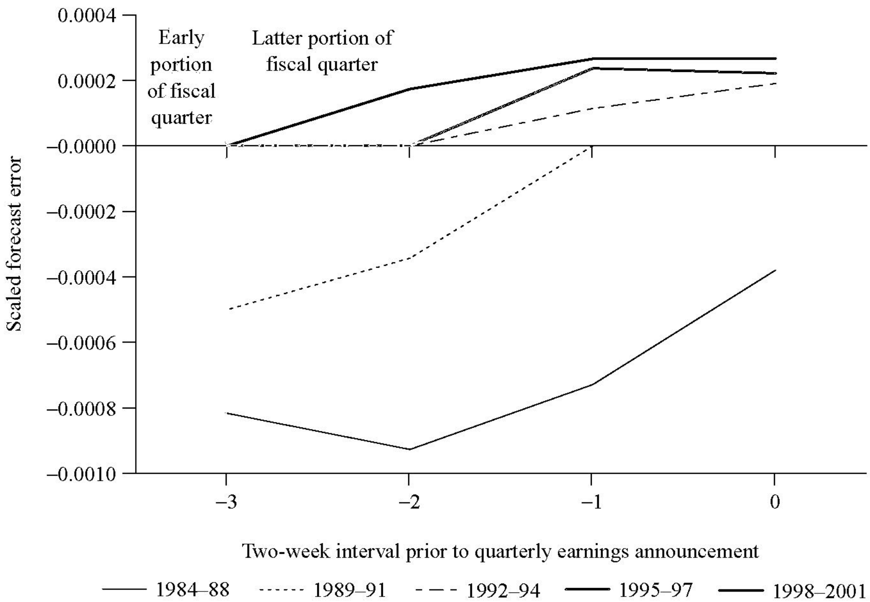
Finally, Baber and Kang (2002) report that forecast errors collected by data providers such as I/B/E/S are rounded to the nearest cent after making retroactive

Figure 1 Median scaled forecast error*

Panel A Annual forecast horizon



Panel B Quarterly forecast horizon



(The figure is continued on the next page.)

Figure 1 (Continued)

Notes:

* The sample includes all firm-year (firm-quarter) observations with data available on the I/B/E/S detail files to construct a median consensus for the monthly (two-week) periods leading up to the annual (quarterly) earnings announcement. All individual analyst forecasts are included except forecasts that create forecast errors greater than stock price (that is, scaled forecasts greater than 100 percent are excluded from the consensus measure). The most recent month (two-week) period prior to the earnings announcement is 0. The sample is broken into five subperiods: 1984–88, 1989–91, 1992–94, 1995–97, and 1998–2001.

and cumulative stock split adjustments. This data-processing artifact compresses analyst forecast errors for firms that have experienced stock splits, which can generate a conservative bias in time-series analyses of forecast errors. Specifically, firms experiencing several stock splits have smaller forecast errors early in times series. The fact that we are still able to document a concentration in small positive forecast errors in recent years speaks to the strength of the walk-down phenomenon. However, as a robustness check, we recalculate our forecast variables using an I/B/E/S data set that does not contain this stock-split problem. Our results are robust using this data set and, therefore, retroactive, and cumulative stock-split adjustments do not explain our results.

In sum, we find evidence of a robust shift toward greater final forecast pessimism. The timing of this shift to pessimism is coincident with the increased use of stock-based compensation in the 1990s and regulatory changes in 1991 concerning the short-swing rule affecting insider's stock option exercises. These changes provide increased managerial incentives to guide analysts to forecast beatable final earnings targets.

5. Quarterly forecast bias and trading incentives

We turn next to tests of the two hypotheses developed in section 2. Although the longer 12-month horizon is useful to show clearly the walk-down pattern over the forecast horizon, we base our tests of the relation between forecast bias and managerial trading incentives using quarterly forecasts.[14] Examining forecasts over the quarterly horizon allows us to focus our analysis on walk-down effects that are not a direct consequence of quarterly earnings announcements. Furthermore, our test results can be compared with recent studies on pessimism in the shortest horizon (e.g., Bagnoli, Beneish, and Watts 1999; Brown 2001; and Matsumoto 2002). Our empirical tests include controls for other factors that affect analyst forecast bias including firm size, growth, and profitability (e.g., Brown 2001).

Table 1 presents descriptive statistics on the sample by calendar subperiods. Firm size is measured at the start of the fiscal quarter as closing stock price at the start of the fiscal quarter (COMPUSTAT data item 14) times the number of common shares outstanding (COMPUSTAT data item 61). The book-to-market ratio is calculated as the book value of common equity at the start of the fiscal quarter

TABLE 1
Descriptive statistics for 53,653 firm-quarter observations for the period 1984–2001

Variable	All years	Year grouping 1984–88	1989–91	1992–94	1995–97	1998–2001
Size ($M)						
Mean	2,571	1,662	1,718	1,758	2,274	4,113
Standard deviation	10,729	3,560	4,701	4,834	7,214	17,638
Q1	137	155	108	127	132	160
Median	422	492	336	376	386	519
Q3	1,504	1,632	1,286	1,302	1,388	1,862
BM						
Mean	0.52	0.596	0.635	0.521	0.473	0.474
Standard deviation	0.38	0.375	0.426	0.324	0.299	0.435
Q1	0.27	0.347	0.346	0.292	0.257	0.217
Median	0.44	0.538	0.552	0.466	0.414	0.383
Q3	0.68	0.771	0.823	0.674	0.621	0.608
Profit Indicator						
Mean	0.87	0.90	0.90	0.90	0.88	0.82
Standard deviation	0.34	0.30	0.30	0.31	0.32	0.38
Q1	1	1	1	1	1	1
Median	1	1	1	1	1	1
Q3	1	1	1	1	1	1
IssueNow						
Mean	0.02	0.015	0.015	0.024	0.020	0.020
Standard deviation	0.06	0.055	0.055	0.073	0.064	0.065
Q1	0	0	0	0	0	0
Median	0.001	0.000	0.001	0.001	0.001	0.002
Q3	0.006	0.004	0.004	0.007	0.006	0.007
IssueNext						
Mean	0.02	0.013	0.013	0.018	0.017	0.018
Standard deviation	0.06	0.047	0.049	0.061	0.056	0.063
Q1	0	0	0	0	0	0
Median	0.001	0.000	0.001	0.001	0.002	0.002
Q3	0.006	0.004	0.004	0.006	0.006	0.007
Insider Sale Indicator						
Mean	0.65	0.666	0.645	0.668	0.682	0.611
Standard deviation	0.48	0.472	0.479	0.471	0.466	0.487
Q1	0	0	0	0	0	0
Median	1	1	1	1	1	1
Q3	1	1	1	1	1	1

(The table is continued on the next page.)

TABLE 1 (Continued)

Variable	All years	Year grouping 1984–88	1989–91	1992–94	1995–97	1998–2001
% Shares Sold						
Mean	0.0014	0.0010	0.0014	0.0016	0.0016	0.0013
Standard deviation	0.0038	0.0030	0.0040	0.0039	0.0040	0.0037
Q1	−0.0000	−0.0000	−0.0000	−0.0000	−0.0000	−0.0001
Median	0.0001	0.0001	0.0001	0.0002	0.0002	0.0001
Q3	0.0013	0.0006	0.0010	0.0014	0.0016	0.0012
Value Shares Sold ($M)						
Mean	1.12	0.46	0.59	0.83	1.16	1.76
Standard deviation	3.39	1.62	1.97	2.44	3.15	4.75
Q1	−0.01	−0.01	−0.01	−0.01	−0.01	−0.02
Median	0.08	0.05	0.05	0.09	0.12	0.91
Q3	0.65	0.31	0.37	0.57	0.83	1.05
Sample size	53,653	6,368	7,098	10,172	14,348	15,667

Notes:

Size is the market capitalization as reported on COMPUSTAT at the start of the fiscal quarter. It is calculated as COMPUSTAT data item 14 (closing stock price at the end of the previous fiscal quarter) multiplied by data item 61 (number of common shares outstanding at the end of the previous quarter).

BM is the book-to-market ratio. It is calculated as the book value of common equity at the start of the fiscal quarter (COMPUSTAT data item 59) divided by market capitalization (Size) at the start of the fiscal quarter.

Profit Indicator is an indicator variable equal to one if EPS as reported on I/B/E/S for the fiscal quarter is positive, and zero otherwise.

IssueNow is the amount of equity issued in the current fiscal quarter. It is calculated as the dollar value of common and preferred equity issued (COMPUSTAT data item 84) divided by market capitalization at the start of the fiscal quarter (that is, at the end of quarter $t-1$).

IssueNext is the amount of equity issued in the next fiscal quarter. It is calculated as the dollar value of common and preferred equity issued (COMPUSTAT data item 84) in quarter $t+1$ divided by market capitalization at the start of quarter $t+1$ (that is, at the end of quarter t).

Insider Sale Indicator is an indicator variable equal to one if the insiders are net sellers of stock in the 20-day period after the quarterly earnings announcement, and zero otherwise. Insiders include the CEO, chair, vice-presidents, officers, and directors. We use the following relationship codes from the Thomson Financial data base: “CB”, “D”, “DO”, “H”, “OD”, “VC”, “AV”, “CEO”, “CFO”, “CI”, “CO”, “CT”, “EVP”, “O”, “OB”, “OP”, “OS”, “OT”, “OX”, “P”, “S”, “SVP”, “VP”.

(The table is continued on the next page.)

TABLE 1 (Continued)

% Shares Sold is the fraction of shares sold by insiders in the 20-day period after the quarterly earnings announcement. This variable is calculated as the net number of shares sold by insiders divided by the number of shares outstanding at the end of the fiscal quarter. The variable is increasing in net sales (that is, negative numbers correspond to net acquisitions by insiders).

Value Shares Sold is the dollar value of shares sold by insiders in the 20-day period after the quarterly earnings announcement. This variable is calculated as the net number of shares sold by insiders multiplied by the price at which those transactions took place. The variable is increasing in net sales (that is, negative numbers correspond to net acquisitions by insiders).

(COMPUSTAT data item 59) divided by market capitalization at the start of the fiscal quarter. Consistent with growth in the economy, the market capitalization has increased and the book-market-to-book ratio has decreased from the 1980s relative to the 1990s. The average value of the profit indicator variable (one if I/B/E/S earnings per share [EPS] for the fiscal quarter are positive, and zero otherwise) shows a marked decline toward the latter half of the 1990s through 2001, consistent with the increase in the number of loss firms over time.[15]

New equity issuance data

One of our key test variables is the firm's own trading activity. We consider two equity issuance variables. *IssueNow* reflects equity issuance in the same quarter as the forecast and *IssueNext* reflects equity issuance in the quarter following the forecast. The issuance variables are measured as the dollar value of common and preferred equity issued from the statement of cash flows (COMPUSTAT data item 84) divided by market capitalization at the beginning of the quarter.[16]

We include *IssueNext* in addition to *IssueNow* because a firm would likely experience similar pressures to avoid an earnings disappointment immediately after issuance. The issuing firm would like to avoid lawsuits from disgruntled investors unhappy with a sizable stock price drop from an earnings disappointment, and the investment banker and analysts of the brokerage firm underwriting the issue would like to safeguard reputation. Table 1 shows a greater level of new equity issuance by firms in the 1992–2001 subperiods relative to the earlier subperiods.

Insider trading data

The second test variable measures managers' trading activity on their personal account. Insider-trading data are obtained from the Thompson Financial insider-trading data base (TFN) covering the period 1984 to 2001. TFN reports all insider trades filed with the SEC resulting from stock transactions and option exercises. We only examine open market sales and purchases of the underlying security (transaction codes "P" and "S" as reported on the data base that originate from Form 4 filings, which include the sale of stock from option exercises). In order to focus on the trading activities of those individuals that are most likely to have an impact on the reporting process of the firm, we include only directors and officers as "insiders" (e.g., the CEO, chair, vice-presidents, and directors) and eliminate trades by nonofficer insiders (e.g., blockholders, retirees, trustees, etc.); see the note in Table 1 for the officer relationship codes. We examine insider trades in the 20 trading days immediately after the earnings announcement.

The *Insider Sale Indicator* equals one if the insiders are net sellers of stock in the 20-day period after the quarterly earnings announcement, and zero otherwise. We also consider two other continuous measures of insider trading activity. *% Shares Sold* is the fraction of shares sold by insiders in the 20-day period after the quarterly earnings announcement. It is the calculated as the net number of shares sold by insiders divided by the number of shares outstanding at the end of the fiscal quarter. The second measure, *Value Shares Sold*, is the dollar value of shares sold by insiders in the 20-day period after the quarterly earnings announcement. This variable is the calculated as the net number of shares sold by insiders multiplied by the price at which those transactions took place. Both continuous measures are increasing in net sales (that is, negative numbers correspond to net acquisitions by insiders).

Table 1 shows a slightly higher frequency of firms with insider selling in the two 1990s subperiods (66.8 percent and 68.2 percent) than in the two subperiods beginning in the 1980s (66.6 percent and 64.5 percent). The lowest frequency of selling (61.1 percent), however, is in the very latest subperiod (1998–2001). A similar pattern is reported for the *% Shares Sold* variable. However, the *Value Shares Sold* variable indicates a monotonic increase over time, perhaps reflecting both the increasing number of stock option exercises as well as increasing stock prices over time.

Cross-sectional variation in forecast bias

Our hypotheses focus on the relation between insider trading behavior and analyst forecast bias. Thus, we group firms by the *Insider Sale Indicator* variable and compare their firm characteristics in Table 2. A firm is classified as a *Seller* in the quarter the *Insider Sale Indicator* equals one, and is classified as a *Purchaser* otherwise. The sample consists of a total of 35,287 *Seller*-quarter and 18,366 *Purchaser*-quarter observations.

Table 2 indicates that *Sellers* are, on average, higher-growth firms as measured by the book-to-market ratios than *Purchasers. Sellers* also are larger firms and more profitable. There is, however, no significant difference in the level of issuing activity.

The key focus of our tests is on the difference between the *Seller* and *Purchaser* groups across samples of firms that differ in the forecast bias in the final month prior to the earnings announcement and in the pattern of analyst forecast bias between long and short horizons. To test Hypothesis 1 directly, we first construct a pessimism indicator variable, $PESS_{last}$, which is equal to one if the price scaled error of the last forecast, $FESC_{last}$, is greater than or equal to zero, and zero otherwise. In other words, the firm was able to meet or beat forecasts in the last month (*Month 0*) prior to the earnings announcement. The Pearson (Spearman) correlation between $PESS_{last}$ and $FESC_{last}$ is 0.48 (0.85). Consistent with analyst guidance incentives associated with insider sales, we find that analysts are significantly more likely to issue pessimistic forecasts for *Seller* firms (66 percent) than for *Purchaser* firms (54 percent).

Next, we calculate a walk-down indicator variable, *SWITCH*, as equal to one if the earliest forecast in the fiscal quarter was optimistic (that is, $FESC_{last} < 0$) and the final forecast in the quarter either equaled actual earnings or was pessimistic (that is, $FESC_{last} \geq 0$), and zero if the first *and* last forecast are both optimistic. This variable is coded as missing for firm-quarter observations where the earliest forecast is pessimistic. Thus, *SWITCH* turns on when the forecast was initially optimistic and the firm was able to meet or beat the forecasts at the end of the quarter. As with the $PESS_{last}$ variable, Table 2 indicates that there is also a significantly higher *SWITCH* for *Sellers* than *Purchasers*, consistent with the prediction in Hypothesis 2.

TABLE 2
Characteristics of firms with net insider sales and net insider purchases following an earnings announcement

Descriptive statistics (means) for firms with insider purchases and insider sales following an earnings announcement. The data set is a pooled time-series cross-sectional sample of 53,653 firm-quarter observations for the period 1984–2001.

	Net insider position		
Variable	Seller, $n = 35{,}287$	Purchaser, $n = 18{,}366$	t-statistic (p-value)
BM	0.458	0.618	−44.09* (<0.001)
MV	6.70	5.89	31.70* (<0.001)
IssueNow	0.0195	0.0194	0.12 (0.90)
IssueNext	0.0163	0.0158	0.92 (0.36)
Profit Dummy	0.90	0.84	17.01* (<0.001)
$PESS_{last}$	0.66	0.54	27.41* (<0.001)
SWITCH	0.27	0.21	11.22* (<0.001)

(The table is continued on the next page.)

TABLE 2 (Continued)

Notes:

A firm is classified as a seller (purchaser) if the insiders are net sellers (purchasers) of company shares in the 20 trading days after an earnings announcement. Insiders include the CEO, chair, vice-presidents, officers, and directors. We use the following relationship codes from the Thomson Financial data base: "CB", "D", "DO", "H", "OD", "VC", "AV", "CEO", "CFO", "CI", "CO", "CT", "EVP", "O", "OB", "OP", "OS", "OT", "OX", "P", "S", "SVP", "VP".

MV is the log of market capitalization as reported on COMPUSTAT at the start of the fiscal quarter. Market capitalization is calculated as COMPUSTAT data item 14 (closing stock price at the end of the previous fiscal quarter) multiplied by data item 61 (number of common shares outstanding at the end of the previous quarter).

BM, *IssueNow*, and *IssueNext* are as defined in Table 1.

Profit Dummy is equal to one if EPS as reported on I/B/E/S for the fiscal quarter is positive, and zero otherwise.

$PESS_{last}$ is an indicator variable equal to one if $FESC_{last}$ is greater than or equal to zero, and zero otherwise. $FESC_{last}$ is the price-scaled median earnings forecast error for analysts covering firm *i*, for earnings in quarter *q*, in the most recent month prior to the quarterly earnings announcement. It is defined as $[Actual\ EPS_{i,q} - Forecast\ EPS_{i,q,t}]/P_{i,q-1}$, where $P_{i,q-1}$ is the stock price when the first forecast is available on I/B/E/S for firm *i* in quarter *q*.

SWITCH is an indicator variable equal to one if the earliest forecast in the fiscal quarter is optimistic (that is, $FESC_{earliest} < 0$) and the final forecast in the quarter is pessimistic (that is, $FESC_{last} \geq 0$), and zero if the first *and* last forecast are both optimistic. This variable is coded as missing for firm-quarter observations where the earliest forecast is pessimistic.

* Significant at the 1% level.

Cross-sectional regression results on forecast pessimism

Table 3 reports the multivariate tests for the cross-sectional determinants of forecast pessimism to evaluate the influence of incentives from insider trading and equity issuance on the final forecast pessimism, after controlling for other factors. We consider two alternative dependent variables, the continuous measure of the scaled forecast error, *FESC*, and the indicator variable for whether the firm beat or met forecast, *PESS*. The measurement of these variables is described above in section 3.

The three key test variables, *InsiderSale*, *IssueNow*, and *IssueNext*, measure the incentives from insider trading and equity issuance. Both *IssueNow* and *IssueNext* are calculated as described earlier. We consider both a binary measure (*InsiderSale Indicator*) as well as a continuous measure for insider selling activity (*%Shares Sold*).[17] These variables are defined above under the heading "Insider trading data". We consider two alternative regression models that differ only in the

TABLE 3

Relation of forecast pessimism with new equity issuance and insider trading

Regression of analyst pessimism on the sale of stock by the firm's CEO in the trading window after the earnings announcement. The data set is a pooled time-series cross-sectional sample of 158,089 firm-quarter-forecast month observations for the period 1986–2001.

Panel A: Scaled forecast error (*FESC*)

$$FESC = \beta_0 + \beta_1{}^*InsiderSale + \beta_2^*IssueNow + \beta_3^*IssueNext + \beta_4^*BM + \beta_5^*MV + \beta_6^*Profit + \beta_7^*Year + \beta_8^*Horizon + \gamma_1^*RD + \gamma_2^*LITIG + \gamma_3^*IMPLICIT + \gamma_4^*CHEARN + \gamma_5^*LABINT + \gamma_6^*LT\_CHEARN + \varepsilon \quad (2b)$$

	Model 1		Model 2	
Variable	*Insider Sale Dummy**	*% Shares Sold**	*Insider Sale Dummy**	*% Shares Sold**
Intercept	−0.016‡	−0.016‡	−0.017‡	−0.017‡
	(−101.4)	(−98.6)	(−94.6)	(−93.1)
InsiderSale	0.002‡	0.147‡	0.001‡	0.096‡
	(32.0)	(20.7)	(23.1)	(13.4)
IssueNow	0.003‡	0.003‡	0.002‡	0.002‡
	(5.94)	(5.65)	(4.11)	(3.85)
IssueNext	0.009‡	0.009‡	0.009‡	0.009‡
	(16.8)	(16.3)	(16.6)	(16.3)
BM	−0.001‡	−0.001‡	−0.0005‡	−0.0006‡
	(−15.8)	(−17.8)	(−6.2)	(−7.5)
MV (*logSize*)	0.0001‡	0.0002‡	0.0002‡	0.0002‡
	(7.5)	(13.6)	(9.8)	(14.1)
Profit	0.013‡	0.013‡	0.012‡	0.012‡
	(158.9)	(158.8)	(132.5)	(132.4)
Year	0.0001‡	0.0002‡	0.0002‡	0.0002‡
	(29.7)	(27.5)	(28.4)	(26.8)
Horizon	0.00054‡	0.0005‡	0.0006‡	0.0006‡
	(19.1)	(18.8)	(20.7)	(20.6)
RD			0.028‡	0.029‡
			(26.8)	(27.3)
LITIG			−0.0005‡	−0.0005‡
			(−8.5)	(−7.6)
IMPLICIT			0.00002‡	0.0001
			(0.3)	(1.72)
CHEARN			0.004‡	0.004‡
			(63.2)	(64.5)
LABINT			−0.0006‡	−0.0006‡
			(−6.4)	(−6.3)
LT_CHEARN			0.015‡	0.015‡
			(29.2)	(29.1)
Model R^2	16.0%	15.7%	19.7%	19.5%
F-value	3,764.7‡	3,677.2‡	2,668.4‡	2,637.1‡

(The table is continued on the next page.)

TABLE 3 (Continued)

Panel B: Pessimism indicator variable (*PESS*)

$$PESS = \beta_0 + \beta_1{}^*InsiderSale + \beta_2^*IssueNow + \beta_3^*IssueNext + \beta_4^*BM + \beta_5^*MV + \beta_6^*Profit + \beta_7^*Year + \beta_8^*Horizon + \gamma_1^*RD + \gamma_2^*LITIG + \gamma_3^*IMPLICIT + \gamma_4^*CHEARN + \gamma_5^*LABINT + \gamma_6^*LT\_CHEARN + \varepsilon \quad (2b)$$

Variable	Model 1 Insider Sale Dummy†	Model 1 % Shares Sold†	Model 2 Insider Sale Dummy†	Model 2 % Shares Sold†
Intercept	−1.64‡	−1.53‡	−2.56‡	−2.51‡
	(2,378.6)	(2,123.2)	(3,818.0)	(3,688.7)
InsiderSale	0.48‡	52.19‡	0.35‡	37.89‡
	(1,751.4)	(1,012.7)	(828.2)	(491.3)
IssueNow	1.10‡	1.05‡	0.87‡	0.82‡
	(113.2)	(102.2)	(60.7)	(54.2)
IssueNext	0.60‡	0.51‡	0.65‡	0.58‡
	(26.8)	(19.1)	(26.9)	(21.5)
BM	−0.17‡	−0.20‡	0.13‡	0.12‡
	(113.5)	(145.5)	(54.9)	(46.7)
MV (*logSize*)	−0.01§	0.02‡	0.02‡	0.05‡
	(4.7)	(49.8)	(37.1)	(157.2)
Profit	1.3266‡	1.32‡	0.92‡	0.92‡
	(5,718.2)	(5,675.9)	(2,137.0)	(2,123.3)
Year	0.0739‡	0.07‡	0.08‡	0.07‡
	(3,244.3)	(2,924.5)	(3,093.3)	(2,889.9)
Horizon	0.18‡	0.17‡	0.21‡	0.21‡
	(925.7)	(898.7)	(1,184.5)	(1,169.4)
RD			4.55‡	4.70‡
			(289.2)	(305.5)
LITIG			0.11‡	0.12‡
			(63.7)	(72.6)
IMPLICIT			0.04§	0.06‡
			(8.3)	(19.8)
CHEARN			1.24‡	1.25‡
			(9,161.6)	(9,352.1)
LABINT			0.18‡	0.17‡
			(74.3)	(69.8)
LT_CHEARN			0.97‡	0.96‡
			(69.8)	(68.5)
Model χ^2	12,257.8‡	11,624.0‡	22,870.0‡	22,567.2‡
p-value	(<0.001)	(<0.001)	(<0.001)	(<0.001)

(The table is continued on the next page.)

TABLE 3 (Continued)

Notes:

Variables are defined as follows:

FESC is the price-scaled median earnings forecast error for analysts covering firm *i*, for fiscal quarter *q* for month *t* prior to the quarterly earnings announcement. It is defined as $(Actual\ EPS_{i,\,q} - Forecast\ EPS_{i,\,q,\,t})/P_{i,\,q-1}$, where $P_{i,\,q-1}$ is the stock price when the first forecast is available on I/B/E/S for firm *i* in quarter *q*.

PESS is an indicator variable equal to one if *FESC* is non-negative, and zero otherwise.

InsiderSale captures the extent of insider trading in the 20-day period following the quarterly earnings announcement. Insiders include the CEO, chair, vice-presidents, officers, and directors. We use the following relationship codes from the Thomson Financial data base: "CB", "D", "DO", "H", "OD", "VC", "AV", "CEO", "CFO", "CI", "CO", "CT", "EVP", "O", "OB", "OP", "OS", "OT", "OX", "P", "S", "SVP", "VP". We use two measures for insider trading. First, we use an indicator variable, *Insider Sale Dummy*. Second, we use a continuous measure, *% Shares Sold*, capturing the fraction of firm traded.

Insider Sale Dummy is an indicator variable equal to one if the insiders are net sellers of stock in the 20-day period after the quarterly earnings announcement, and zero otherwise.

% Shares Sold, *IssueNow*, *IssueNext*, and *BM* are as defined in Table 1.

MV is as defined in Table 2.

Profit is an indicator variable equal to one if EPS as reported on I/B/E/S for the fiscal quarter is positive, and zero otherwise.

Year captures the time trend in forecast errors. It is the year in which the forecast is made less 1984 (the first year in the sample).

Horizon captures the time between the forecast and the earnings announcement. It is calculated as the number of months prior to the quarterly earnings announcement. For example, a forecast made in February (April) for a fiscal quarter ending March 31 with an announcement date of April 14 corresponds to a value of −2 (0) for *Horizon*. *Horizon* is increasing in closeness to the earnings announcement.

RD is research and development expenditure (COMPUSTAT data item 4). It is scaled by average total assets (COMPUSTAT data item 44).

LITIG is an indicator variable equal to one for high litigation risk industries as defined by Matsumoto (2002), and zero otherwise. The industry four-digit SIC codes for high litigation industries include 2833, 2836, 3570, 3577, 3600–3674, 5200–5961, and 7370–7374.

IMPLICIT is an indicator variable equal to one for industries with a high degree of reliance on implicit claims by stakeholders as defined by Matsumoto 2002, and zero otherwise. The industry four-digit SIC codes for these industries include 150–179, 245, 250–259, 283, 301, 324–399.

(The table is continued on the next page.)

TABLE 3 (Continued)

Notes:

CHEARN is an indicator variable equal to one for a positive change in earnings from the same quarter in the prior year (COMPUSTAT data item 8), and zero otherwise. This variable is the same as in Matsumoto 2002.

LABINT is a measure of labor intensity. It is calculated as [1 – (*PPE*/*Gross Assets*)]. *PPE* is property, plant, and equipment (COMPUSTAT data item 118). *Gross Assets* is calculated as the sum of total assets (COMPUSTAT data item 44) and accumulated depreciation and amortization (COMPUSTAT data item 41). See also Matsumoto.

LT_CHEARN is a measure of long-term change in earnings. It is the change in earnings from four quarters prior to the forecast quarter to four quarters after the forecast quarter. The measure is scaled by the market capitalization of the firm four quarters prior to the forecast quarter.

* *t*-statistics are reported in parentheses.

† χ^2 statistics are reported in parentheses below parameter estimates.

‡ Significant at the 1 percent level.

§ Significant at the 5 percent level.

set of control variables. The inclusion of these variables helps evaluate the incremental influence of insider trading and equity issuance incentives beyond the other incentives identified by Matsumoto 2002. The first regression model is

$$FESC \text{ or } PESS = \beta_0 + \beta_1 InsiderSale + \beta_2 IssueNow + \beta_3 IssueNext + \beta_4 BM + \beta_5 MV + \beta_6 Profit + \beta_7 Year + \beta_8 Horizon + \varepsilon \quad (2a).$$

Drawing from previous research (e.g., Brown 2001 and Matsumoto 2002), the control variables in model 1 include firm size, growth, and profitability. *Profit* is an indicator variable equal to one if EPS as reported on I/B/E/S for the fiscal quarter is positive, and zero otherwise. *MV* is the log of market capitalization as reported on COMPUSTAT at the start of the fiscal quarter (defined earlier). Because a high-growth firm would likely need new capital, and would also care about investor perceptions and want to avoid an earnings disappointment, we include a growth proxy, *BM*. It is calculated as the book value of common equity at the start of the fiscal quarter divided by market capitalization (*MV*) at the start of the fiscal quarter.

We use a pooled time-series cross-sectional regression framework, so we also include two additional variables to pick up possible changes in forecast pessimism over the calendar time as well as over the forecast horizon. *Year* captures the calendar time trend in forecast errors and is measured by the difference between the calendar year of the forecast and the base year 1984 (the first year in the sample). *Horizon* captures the time between the forecast and the earnings announcement. It is calculated as the number of months prior to the quarterly earnings announcement. For example, a forecast made in February (April) for a fiscal quarter ending March 31 with an announcement date of April 14 corresponds to a value of −2 (0) for *Horizon*. *Horizon* is increasing in closeness to the earnings announcement.

The second regression model is

$$\begin{aligned} FESC \text{ or } PESS = {} & \beta_0 + \beta_1{}^{*}InsiderSale + \beta_2{}^{*}IssueNow + \beta_3{}^{*}IssueNext + \beta_4{}^{*}BM \\ & + \beta_5{}^{*}MV + \beta_6{}^{*}Profit + \beta_7{}^{*}Year + \beta_8{}^{*}Horizon + \gamma_1{}^{*}RD \\ & + \gamma_2^{*}LITIG + \gamma_3^{*}IMPLICIT + \gamma_4^{*}CHEARN + \gamma_5^{*}LABINT \\ & + \gamma_6^{*}LT\_CHEARN + \varepsilon \qquad (2b). \end{aligned}$$

In addition to the control variables in the first model, model 2 includes proxies for a firm's litigation risk, reliance of financial information by noninvestor stakeholders, and further proxies for a firm's future profitability prospects. Sivakumar and Vijaykumar (2001) and Matsumoto (2002) suggest that these factors affect a firm's ability to meet or beat forecasts.

We use an indicator variable, *LITIG*, equal to one for high litigation risk industries as defined by Matsumoto 2002, and zero otherwise; see notes to Table 3 for the four-digit SIC codes considered to be high litigation risk industries. We also use the three Matsumoto variables to control for the effects on forecast pessimism that is derived from a greater reliance of financial information for implicit claims by non-investor groups. *RD* is research and development expenditure (COMPUSTAT data item 4) scaled by average total assets (COMPUSTAT data item 44). *IMPLICIT* is an indicator variable equal to one for the durable goods industries, and zero otherwise; see notes to Table 3 for the four-digit SIC codes. *LABINT*, a measure of labor intensity, is calculated as [1 − (*PPE*/*Gross Assets*)] where *PPE* is property, plant, and equipment (COMPUSTAT data item 118), and *Gross Assets* is the sum of total assets (COMPUSTAT data item 44) and accumulated depreciation and amortization (COMPUSTAT data item 41).

The final two control variables are related to the firm's current and future profitability. *CHEARN*, is an indicator variable equal to one for a positive change in earnings (COMPUSTAT data item 8) from the same quarter in the prior year, and zero otherwise. This controls for possible contemporaneous unexpected shocks to earnings that may affect the firm's ability to meet or beat forecasts independent of the strategic behavior by the firm to guide forecasts.

LT_CHEARN is calculated as the change in earnings from four quarters prior to the forecast quarter to four quarters after the forecast quarter, scaled by the market capitalization of the firm four quarters prior to the forecast quarter. The long-term change in earnings, suggested by Sivakumar and Vijaykumar 2001, controls for the possibility that the firm's long-term prospects may influence the manager's trading behavior on the firm's or the manager's own behalf, as well as the firm's ability to beat or meet current forecasts.

The ordinary least squares (OLS) pooled cross-sectional regression is run when *FESC* is the dependent variable, and a logistic regression is run when *PESS* is the dependent variable.[18] The results reported in Table 3 are consistent with the predictions of Hypothesis 1. The three key test variables *InsiderSale*, *IssueNow*, and *IssueNext* are all highly statistically significant in the predicted direction, confirming that managerial and firm incentives to sell equity are significantly associated with whether firms meet or beat forecasts.

Taking *InsiderSale* first, Table 3 reports that greater forecast pessimism is found for firms with higher insider selling subsequent to the quarter when they beat or meet the quarterly consensus earnings forecast. In panel A, all else constant, a firm that had net insider selling after the earnings announcement and an average price–earnings (P/E) ratio of 30 would beat forecasts by an average of 5.34 percent (estimated coefficient for *InsiderSale* \$0.00178*30) more than a firm that had net insider purchase. A similar message is obtained when the dependent variable is an indicator variable of whether the firm beat or met forecasts.

The analysis in the first column of Table 3 (panel B) reports that the log odds ratio of beating or meeting increases by 48 percent when insiders are net sellers in the 20-day window following the earnings announcement. Alternatively stated, the probability of a pessimistic forecast error is 21 percent higher for a firm with net insider selling compared with a firm with net insider purchases (calculated using mean values for independent variables in the model 1 regression). The result of a positive association between forecast pessimism and insider selling is robust when insider selling is measured as a percentage of shares sold, and is also robust to the set of control variables included.

Turning to the equity issuance incentives, Table 3 reports that *IssueNow* and *IssueNext* representing equity issuance in the same quarter and in the future quarter respectively are associated with positive earnings surprises. For example, in panel A, a firm with an average P/E of 30 that issued an additional 10 percent of its market value in the quarter following the earnings announcement, on average, beat forecasts by about 2.8 percent (\$0.00929*0.1*30) more than a firm that did not issue new equity. In panel B, a firm that issues an additional 10 percent of its market value in the subsequent quarter experiences a 3 percent higher probability of beating or meeting forecasts than a firm that did not issue new equity (calculated as the marginal probability increase for an additional 10 percent of new equity in the following quarter, holding all variables at their mean values). As for *InsiderSale*, the results for the issuance variables are also robust with respect to the set of control variables included in the regression.

Furthermore, the evidence for quarterly forecasts in Table 3 further corroborates the pattern of annual forecast errors, consistent with a forecast walk-down illustrated in Figure 1. The significantly positive *Horizon* coefficient indicates that forecast pessimism increases as the forecast horizon shrinks toward the earnings announcement, consistent with a walk-down in forecasts. The significantly positive *Year* coefficient indicates that forecast pessimism has increased with calendar time from the 1980s to 2001.[19]

The results reported above are robust with respect to whether the measures of pessimism and insider selling are continuous or binary (*FESC* or *PESS*; *InsiderSale* or *% Shares Sold*), and whether a partial or full set of control variables is included in the regression. The first set of control variables includes firm size, growth opportunities, and profitability. Not surprisingly, ex post profitable firms tend to beat analysts' targets because the earnings realization turned out to be high. Similarly, growth firms as proxied by low book-to-market ratios also demonstrate a greater likelihood of the firm beating or meeting forecasts. With one exception, the results for firm size suggest that larger firms are more able to meet or beat forecasts.

Our results for the additional control variables are consistent with the findings in past studies. Consistent with Matsumoto (2002), the model 2 regression results in Table 3 indicate that firms with high litigation risk or a high reliance on implicit claims with stakeholders are more likely to meet or beat forecasts. Consistent with Sivakumar and Vijaykumar 2001, firms with past long-term growth in earnings are also more able to beat or meet forecasts. Consistent with the managerial guidance hypotheses, our key results here indicate that the equity-issuance and managerial insider-selling incentives exert an incremental influence on forecast pessimism over these additional explanatory variables.

The cross-sectional regressions presented in Table 3 are estimated using a pooled sample from 1984–2001 (some 158,089 firm-quarter-month observations). To examine the impact of forecast horizon, our pooled sample includes multiple firm observations for each firm-quarter. This may raise a concern of dependence in the data. Specifically, we have up to three observations for each firm-quarter. The inclusion of the fixed effects horizon variable may only partially address this dependence. Therefore, as an additional robustness check on the regression specification, we run regressions using only one (the final) forecast for each firm-quarter. We exclude the horizon variable from this specification (as we have only one record per firm-quarter). The results from this reduced sample of 53,653 firm-quarter observations yield similar results. With the exception of the *IssueNow* variable, which loses significance after inclusion of the Matsumoto 2002 control variables, we continue to find strong statistical (t-statistics range between 6.47 and 16.55 for the alternative specifications) and economic significance for *IssueNext* and the insider selling variable (both the indicator and continuous variables) in both the *FESC* and *PESS* regressions.

As a final sensitivity check, we also perform 60 quarterly cross-sectional regressions for the *FESC* dependent variable to obtain Fama-Macbeth 1973 t-statistics calculated from the time series of the estimated quarterly cross-sectional regression coefficients; results are not tabulated. *Year* and *Horizon* variables are not included in this specification. We include the three control variables for firm size, growth opportunities, and profitability. Both insider-selling variables remain highly statistically significant (t-statistics of 10.31 for the indicator variable and 5.70 for the continuous variable). The *IssueNow* and *IssueNext* variables are marginally significant in these specifications (t-statistics of between 1.72 and 1.96). The lower statistical significance from the Fama-Macbeth procedure reflects the lower power from equally weighting the time-series observations (e.g., Loughran and Ritter 2000).

Determinants of the switch from initial forecast optimism to final pessimism

The empirical findings reported in the previous section are consistent with the predictions of Hypothesis 1. However, we are careful to note that the observed association between pessimistic analyst forecast revisions and our trading measures may also be consistent with managers' ex post timing equity sales when price is relatively high (after truly unexpected good earnings). However, the univariate tests reported in Table 2 indicate that *Sellers* are more likely to experience a switch from forecast optimism to pessimism during the quarter than *Purchasers*. This switching behavior seems more consistent with opportunistic guidance. Therefore, to test the more restrictive predictions of Hypothesis 2, we estimate logistic cross-sectional regressions of the *Switch* indicator variable (described under the heading "Cross-sectional variation in forecast bias") using the key test variables and the same set of control variables as in Table 3 regressions.

$$\begin{aligned} SWITCH = {} & \beta_0 + \beta_1{}^*InsiderSale + \beta_2{}^*IssueNow + \beta_3{}^*IssueNext + \beta_4{}^*MB \\ & + \beta_5{}^*MV + \beta_6{}^*Profit + \beta_7{}^*Year + \gamma_1{}^*RD + \gamma_2{}^*LITIG + \gamma_3{}^*IMPLICIT \\ & + \gamma_4{}^*CHEARN + \gamma_5{}^*LABINT + \gamma_6{}^*LT\_CHEARN + \varepsilon \qquad (3). \end{aligned}$$

Given the definition of the *Switch* variable, the estimation of (3) is restricted to the sample of firms where the forecasts are initially optimistic.[20] The results are reported in Table 4. As in Table 3, *InsiderSale* in Table 4 is highly statistically significant, which is consistent with insiders timing their sales to follow immediately after a good news earnings surprise, and consequently after an increase in stock price. Relative to *Purchaser* firms, *Seller* firms experience a 21 percent higher probability of a switch from early optimism to final pessimism (calculated as the probability difference from comparing firms with net insider sales to firms with no net insider selling, holding all other variables at their mean values). Similarly, *IssueNow* and *IssueNext* are also highly statistically significant in model 1 regressions. An equity issuance equal to 10 percent of market capitalization in the subsequent quarter is associated with a 6 percent higher probability of a switch in early optimism to final pessimism, compared with a firm with no equity issuance in the following quarter. Although *IssueNext* remains highly significant in model 2 regressions, *IssueNow* does not, perhaps because of high correlation with the additional included variables. These results support the predictions of Hypothesis 2.

The statistically significant result for *Year* indicates that there is a greater likelihood of a switch from initial optimism to final pessimism in more recent calendar years, further confirming the predictions of Hypothesis 2. Institutional changes during the 1990s increased the firm's economic incentives to walk-down forecasts and then to beat or meet them at the earnings-announcement date.

The control variables have similar effects on the *SWITCH* indicator as on the *PESS* indicator described in Table 3. Larger firms that have more growth opportunities and that are profitable are more likely to have forecasts that switched from being optimistic to pessimistic over the forecast horizon. Finally, some of the implicit claims and litigation risk proxies are significant (*LITIG, IMPLICIT, CHEARN*), but others are not (*RD, LABINT, LT_CHEARN*).

TABLE 4

Relation of switching from initial optimism to final pessimism with new equity issuance and insider trading

Regression of a switch from forecast optimism to pessimism, on the sale of stock by the firm's CEO in the trading window after the earnings announcement. The data set is a pooled time-series cross-sectional sample of 25,414 firm-quarter observations for the period 1984–2001.

$$SWITCH = \beta_0 + \beta_1{}^*InsiderSale + \beta_2{}^*IssueNow + \beta_3{}^*IssueNext + \beta_4{}^*MB + \beta_5^*MV + \beta_6^*Profit + \beta_7{}^*Year + \gamma_1{}^*RD + \gamma_2^*LITIG + \gamma_3^*IMPLICIT + \gamma_4^*CHEARN + \gamma_5^*LABINT + \gamma_6^*LT\_CHEARN + \varepsilon \quad (3)$$

	Model 1		Model 2	
Variable	*Insider Sale Dummy**	*% Shares Sold**	*Insider Sale Dummy**	*% Shares Sold**
Intercept	−3.18†	−3.02†	−3.48†	−3.43†
	(1,142.3)	(1,112.4)	(990.5)	(973.0)
InsiderSale	0.25†	25.37†	0.21†	20.28†
	(62.0)	(33.3)	(40.0)	(19.5)
IssueNow	0.77†	0.78†	0.65‡	0.65‡
	(7.0)	(7.2)	(4.6)	(4.6)
IssueNext	0.81†	0.75‡	0.92†	0.88†
	(6.7)	(5.7)	(7.7)	(7.0)
BM	−0.30†	−0.32†	−0.16†	−0.17†
	(35.8)	(40.2)	(8.9)	(10.3)
MV (*logSize*)	0.10†	0.11†	0.10†	0.12†
	(103.5)	(138.2)	(112.8)	(142.3)
Profit	0.89†	0.89†	0.81†	0.81†
	(334.6)	(331.8)	(235.1)	(233.5)
Year	0.06†	0.06†	0.07†	0.06†
	(300.5)	(279.4)	(303.4)	(287.3)
RD			0.71	0.83
			(1.1)	(1.5)
LITIG			0.18†	0.18†
			(23.5)	(24.5)
IMPLICIT			0.12†	0.13†
			(12.0)	(14.5)
CHEARN			0.36†	0.37†
			(112.7)	(118.8)
LABINT			−0.06	−0.06
			(1.2)	(1.2)
LT_CHEARN			−0.26	−0.26
			(0.6)	(0.6)
Model χ^2	1,167.7†	1,136.1†	1,308.2†	1,286.8†
p-value	(<0.001)	(<0.001)	(<0.001)	(<0.001)

(The table is continued on the next page.)

TABLE 4 (Continued)

Notes:

This table uses only one observation for each firm-quarter. Therefore, the horizon variable is dropped from the analysis.

Variables are defined as follows:

InsiderSale captures the extent of insider trading in the 20-day period following the quarterly earnings announcement. This is measured using an indicator variable, *Insider Sale Dummy* (equal to one if the insiders are net sellers of stock in the 20-day period after the quarterly earnings announcement, and zero otherwise), or a continuous measure, *% Shares Sold* (the fraction of shares sold by insiders in the 20-day period after the quarterly earnings announcement). This variable is calculated as the net number of shares sold by insiders divided by the number of shares outstanding at the end of the fiscal quarter. The variable is increasing in net sales (that is, negative numbers correspond to net acquisitions by insiders). Insiders include the CEO, chair, vice-presidents, officers, and directors. We use the following relationship codes from the Thomson Financial data base: "CB", "D", "O", "H", "OD", "VC", "AV", "CEO", "CFO", "CI", "CO", "CT", "EVP", "O", "OB", "OP", "OS", "OT", "OX", "P", "S", "SVP", "VP".

IssueNew, *IssueNext*, and *BM* are as defined in Table 1.

Switch and *MV* are as defined in Table 2.

All other variables are as defined in Table 3.

* χ^2 statistics are reported in parentheses below parameter estimates.

† Significant at the 1 percent level.

‡ Significant at the 5 percent level.

In unreported tests, we find similar, if not stronger, results using annual forecast horizons in documenting the relation between equity issuance/insider selling and forecast pessimism and the switch from forecast optimism to pessimism. Taken together, the results from Tables 2, 3, and 4 are consistent with managers guiding analyst earnings targets to facilitate trading on favorable terms after an earnings announcement, on both the manager's and the firm's behalf. The potential for the manager or firm to benefit from these transactions is derived from the managers' ability to guide analysts over the forecast horizon prior to trading.

Robustness analysis and discussion of limitations

In this section, we report two additional robustness checks and discuss some caveats concerning the interpretation of our results. The first robustness check examines whether analyst pessimism varies with analyst type. If bias differs across analysts, then firm variation in a forecast walk-down could result from the presence of different analyst types rather than from varying incentives of managers and firms to sell stock after the earnings announcement.

We compare the forecast errors and forecast pessimism between "lead" and "follower" analysts, where "lead" and "follower" types are identified using an approach analogous to Cooper, Day and Lewis (2001). Similar to Cooper et al., we ignore forecasts in the first 30 days of the quarter and focus instead on analyst forecasts issued in the last 30 days of the quarter, which are more likely to be revisions resulting from unobservable managerial guidance. Analysts who revise their earnings forecast first in the last 30 days of the quarter are identified as "lead" analysts. To ensure that a "lead" analyst is truly a first mover, we require a 10-day quiet window preceding forecast revision of the "lead" analyst. If multiple analysts revise their forecasts on the same day, the value of the "lead" forecast is calculated as the mean of the analyst forecasts issued on that day. "Follower" analysts are identified as those analysts who revise their forecasts in the days following the "lead" analysts, but before the actual earnings announcement. The sample consists of 12,157 firm-quarter observations.

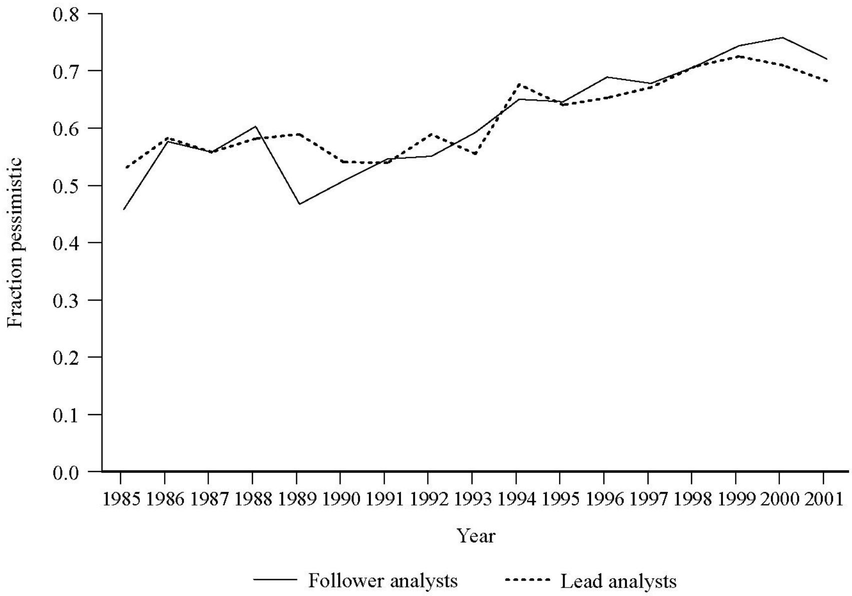
Our empirical results show no economic or statistical difference between the forecast bias properties of "lead" analysts and those of "follower" analysts. For example, the average pessimism ($PESS_{last}$) for "lead" analysts is 0.644 over the entire sample period while the average pessimism for "follower" analysts is nearly identical at 0.638, and the difference is not statistically significant. Figure 2 presents the temporal trend of pessimism in "lead" and "follower" analyst forecast revisions for the period 1985–2001. The graph shows increasing pessimism for both "lead" and "follower" analysts over the sample period, similar to the graph for the consensus forecasts in Figure 1. There is, however, no statistical difference between the two categories of analysts.

These findings are consistent with the notion that managers have strong incentives to manage the consensus of all analysts' earnings forecasts. While it may be important to first guide influential "lead" analysts, managers must ultimately guide the consensus of all analyst forecasts because the consensus earnings estimate is the benchmark used to evaluate subsequent reported earnings. Furthermore, the statistically indistinguishable difference between forecasts of lead and follower analysts is consistent with the analyst herding behavior reported in prior studies (see, for example, Hirshleifer and Teoh, 2003).

Our second robustness check examines the impact of different investor types — namely, institutional versus noninstitutional investors — on analyst forecast bias. We reestimate our main regressions using a subsample (140,906 firm-quarter-forecast month observations) with institutional holdings data available from the 2001 Spectrum data base. These regressions now include a variable measuring the fraction of shares held by institutional investors. Our main findings on the relation between insider sales and analyst forecast errors and pessimism remain robust for this subsample. Consistent with Matsumoto 2002, we also find a positive association between the fraction of institutional ownership and forecast pessimism. This finding is consistent with the argument that the increasingly short-term investment objectives of institutional investors may provide managers with additional pressures to beat short-term quarterly targets. The descriptive findings of Matsumoto also suggest that the effect is strongest for transient institutional investors.

While our empirical results are robust to a number of different specifications, as in all empirical research, caution is required in interpreting the findings. The focus of this paper is to identify determinants of (1) forecast pessimism at the end of the fiscal year, and (2) the switch from early optimism to final pessimism. In developing our hypotheses, we rely on the prior research of Bartov, Givoly, and Hayn 2002 to support our premise that analyst guidance leads to more favorable stock prices at the end of the fiscal period. This prior evidence suggests that the path by which forecasts come to be beaten is not as crucial as whether the forecast is beaten. Our finding that final pessimism and the switch from early optimism to final pessimism is concentrated in firms that are net issuers of equity or managers are net sellers of stock after an earnings announcement is consistent with these firms choosing to engage in such behavior because of managerial incentives. Therefore, our results should be interpreted as a joint test of (1) the hypothesis that the forecast path is less crucial than whether the forecast is beaten, and (2) our earnings-guidance hypothesis.

Figure 2 Temporal trend of pessimistic lead and follower analysts*



Notes:

* To identify lead and following analysts we use a procedure similar to Cooper, Day, and Lewis 2001. We focus on analysts releasing forecasts in the last month of the fiscal quarter and require there be no forecasts in the first third of the last month (that is, days −30 to −21) to ensure there is no significant news event. We then divide the forecasts made in the last 20 days into the first forecast (lead analyst) and take the average of the remaining forecasts (followers).

In this paper, we investigate expectations management as one of several tools that management has available to achieve a desired level of earnings-surprise. It should be noted that our earnings-surprise measure compares analysts' earnings estimates with a firm's reported earnings. The reported earnings number can also be managed (for example, by manipulating accruals or changing earnings definitions) to achieve the desired earnings surprise (e.g., Teoh, Welch, and Wong 1998a, 1998b; and Bradshaw and Sloan 2002). Therefore, we view our results as providing complementary (and often inseparable) evidence on both earnings and expectations management.

Several recent U.S. regulatory reforms may limit the ability of analysts and managers to engage in future earnings guidance games. The enactment of Regulation FD (Fair Disclosure), in October 2000, may limit managers' hidden opportunities to guide analysts' forecasts. In addition, the enactment of Regulation AC (Analyst Certification) in 2003 requires analysts to certify that recommendations reflect their personal beliefs. However, to the extent that none of the current regulations require firms to disclose at the time of the earnings announcement the firm's or insiders' intention to sell the firm's stock shortly after the earnings announcement, these economic incentives may still be present to encourage continuation of the earnings-guidance game.

6. Conclusion

This paper examines the dynamic behavior of analyst earnings forecasts leading up to earnings announcements. We provide evidence that links the pattern of analyst pessimism in the 1990s to institutional and regulatory changes that create capital-market incentives for managers to guide and beat forecasts in order to boost stock prices. These systematic changes include greater use of stock option compensation for managers, restrictions on trading by insiders to post-earnings-announcement periods in response to the Insider Trading and Securities Fraud Enforcement Act of 1988, and the lifting of the short-swing rule for insiders in 1991 allowing insiders to exercise stock options and immediately sell company stock.

Our cross-sectional predictions are motivated by the tendency of managers and firms to sell shares after earnings announcements. This can create incentives to guide analysts to systematically pessimistic forecasts just prior to the earnings announcement, so that the salient news of a positive rather than a negative surprise arrives before the share sale.

Consistent with our hypotheses, we find that pre-announcement forecast pessimism is strongest in firms whose managers have the highest capital-market incentives to avoid earnings disappointments. We find that firms with managers that sell stock after an earnings announcement are more likely to have pessimistic analyst forecasts prior to the earnings announcement. The probability of forecast pessimism increases from 54 percent for an average firm without net insider selling to 66 percent for an average firm with subsequent net insider selling. Furthermore, firms in which the insiders are net sellers of the firm's stock are also more likely to have analysts switch from long-horizon optimism to short-horizon pessimism prior to the earnings announcement. The probability of a switch from optimism early in the quarter to pessimism closest to the earnings announcement increases from 21 percent in firms without net insider selling to 27 percent in firms with net insider selling.[21] This evidence is consistent with managers behaving opportunistically to guide analysts' expectations around earnings announcements to facilitate favorable insider trades after earnings announcements.

Endnotes

1. Cotter, Tuna, and Wysocki (2004) examine analysts' forecast revisions in response to public managerial guidance as provided through management's earnings forecasts. However, prior to Regulation FD (SEC 2000), a large fraction of managerial guidance of analysts was not publicly observable.
2. For example, one might speculate that managers are just opportunistically taking advantage of unrelated changes in analyst forecast bias by selling shares or exercising options. However, we are not aware of any specific explanation for why their incentive to do so would cause them to behave in a way that explains our evidence.
3. Managers also care about the stock price performance because poor stock price performance encourages a hostile takeover and subsequent firing by the acquirer's board of directors. An active external labor market also rewards a manager with a reputation for maintaining good stock price performance. In addition, a manager is in a better position to bargain for higher future compensation if the stock price performance is good.
4. By reducing discretion in the timing of the insider trades, the blackout feature reduces the opportunity of the managers to profit from inside information at the expense of uninformed outside investors. Limiting insider trades to the period immediately after earnings announcements also reduces the adverse selection problem by minimizing the asymmetry of information between uninformed outsiders and the inside managers.
5. See http://www.sec.gov/rules/final/33-8193.htm for full details. Part A of the Final Rule indicates the following:

> A. Certifications in Connection with Research Reports: As adopted, Regulation Analyst Certification requires that brokers, dealers, and their associated persons that are "covered persons" that publish, circulate, or provide research reports include in those research reports:
>
> (A) a statement by the research analyst (or analysts) certifying that the views expressed in the research report accurately reflect such research analyst's personal views about the subject securities and issuers; and
>
> (B) a statement by the research analyst (or analysts) certifying either:
>
> (1) that no part of his or her compensation was, is, or will be directly or indirectly related to the specific recommendations or views contained in the research report; or
>
> (2) that part or all of his or her compensation was, is, or will be directly or indirectly related to the specific recommendations or views contained in the research report. If the analyst's compensation was, is, or will be directly or indirectly related to the specific recommendations or views contained in the research report, the statement must include the source, amount, and purpose of such compensation, and further disclose that it may influence the recommendation in the research report.

6. This does not require that investors be irrational in their evaluations of forecasts. Investors may properly discount for optimism, but firms nevertheless need to induce such analyst optimism because investors would still discount a defecting firm that failed to do so, causing that firm to be viewed as worse than it really is.
7. The increased use of stock options in the 1990s may have been, in part, an endogenous favorable response by firms to the reduced agency-related costs of stock option compensation that resulted from the heightened insider-trading restrictions (discussed above under the heading "Why and when managers care about short-term stock prices"). The findings in this study suggest that we may have substituted one agency-related cost for another. The new agency cost is one that resulted from an increased incentive to play the earnings-guidance game.
8. It is important to note that our analysis of the switch from early optimistic to pessimistic forecasts does not collapse to an analysis of final pessimism. In considering the optimism–pessimism switch we exclude firm-quarter observations where the initial forecast is pessimistic. More details on variable measurement are given in section 5.
9. Our results are not driven by use of this "constructed" consensus forecast. In unreported tests we replicate our empirical analysis using the median consensus forecast as reported by I/B/E/S.
10. The empirical findings documented in this section also exist for a broader sample of firms not restricted by COMPUSTAT data availability.
11. We also replicate the analysis using total assets per share as a deflator. The qualitative results are unchanged using this alternative deflator.
12. For example, an analyst forecasts $1.15 earnings per share (EPS) for a firm on November 1, 1995 for the fiscal year ending December 31, 1995. I/B/E/S reports an actual EPS of $1.20 on January 27, 1996. I/B/E/S also reports that the 1994 fiscal year earnings release date occurs during January 1995, and the stock price in February 1995 (the first month after the release of EPS for the previous fiscal year) is $15.10. Thus, *FE* for month –2 (73 days' lag between earnings release date and forecast date) is ($1.20 – $1.15)/$15.10 = 0.0033, or 0.33 percent. We use a calendar-year timing convention, so the *FE* is considered the forecast error for year 1996 because the actual earnings release date occurs in January 1996.
13. For example, absolute forecast errors (|forecast EPS – actual EPS|) greater than $3 per share for a company trading at $30 per share are removed from the sample. Data-coding errors for forecasts and extreme small prices likely contribute to such large outliers. The 10 percent deletion rule removed 2.1 percent of the sample. We find that the mean (median) numerator of *FESC* is –0.04 (0.00) for retained firms and –1.20 (–0.66) for deleted firms. Further, we find that the mean (median) denominator of *FESC* is 28.76 (19.25) for retained firms and 5.73 (3.50) for deleted firms. Deleted firms have much larger unscaled forecast errors and lower stock prices. As a robustness check, we apply a less stringent deletion cutoff of greater than 100 percent of price that removes only 0.2 percent of the sample. Our results are qualitatively unchanged in this specification and remain statistically significant.
14. Our empirical findings are stronger in tests (not reported) using annual horizons.
15. Givoly and Hayn (2000) report a loss frequency of about 34 percent in the 1990s based on net income. Our sample is skewed toward larger (more profitable) firms with analyst

following. In addition, we use I/B/E/S income numbers, which are typically based on operating earnings.

16. The empirical results are robust to the use of an equity-issuance indicator variable based on equity-sale cutoffs from 1 percent to 20 percent of equity market value. For the indicator variables, we exclude the smallest equity issuances because they relate to additional equity issued due to the exercise of managerial options. For the continuous variables, we note that the issuance variable may be correlated with the insider trade variable via stock options exercise. The Pearson (Spearman) correlation between the insider selling and equity-issuance variables is 0.18 (0.21).
17. Regression results for the second continuous measure of insider trading (dollar value of shares traded) are similar to the fraction of shares traded variable. We do not report these results for the sake of brevity.
18. In additional tests we also considered the robustness of the regression results in panel B of Table 3 to our definition of *PESS*. If we limit our categorization of firms who meet/beat (miss) to those firms who report earnings no more than 5 cents greater (lower) than the most recent consensus analyst estimate all of our explanatory variables retain their significance. This reinforces the earlier discussion that firms need only *just* beat analyst expectations. Managerial incentives to sell equity both on the firm's behalf and from their own personal accounts are a key determinant in the discontinuity of analyst forecast errors around the zero point.
19. In unreported tests, we also interact the equity-issuance and growth variables with the temporal trend. There is some indication that these effects are more pronounced in the latter part of our sample. In addition, our findings are robust to the inclusion of annual and quarterly fixed effect variables.
20. We reran the analysis in Table 3 using this restricted sample where the initial forecasts are optimistic. The results are essentially the same, and the key variables related to our hypotheses remain statistically significant using the reduced sample.
21. Although the economic magnitude of these quarterly forecast results is modest, the annual forecast results are more substantial. This is because there is a much larger fraction of optimistic forecasts at the beginning of the fiscal year (> 70 percent) than at the start of a fiscal quarter (< 50 percent); this difference has increased in the latter years in our sample period as firms appear to walk-down forecasts to beatable levels earlier and earlier in the fiscal period.

References

Abarbanell, J., and V. Bernard. 1992. Tests of analysts' overreaction/underreaction to earnings information as an explanation for anomalous stock price behavior. *Journal of Finance* 47 (3): 1181–207.

Aboody, D., and R. Kasznik. 2000. CEO stock option awards and corporate voluntary disclosures. *Journal of Accounting and Economics* 29 (1): 73–100.

Baber, W. R., and S.-H. Kang. 2002. The impact of split adjusting and rounding on analysts forecast error calculations. *Accounting Horizons* 16 (4): 277–89.

Bagnoli, M., M. Beneish, and S. Watts. 1999. Whispers and shouts: Forecasts of quarterly earnings per share. *Journal of Accounting and Economics* 28 (1): 27–50.

Bartov, E., D. Givoly, and C. Hayn. 2002. The rewards to meeting-or-beating earnings expectations. *Journal of Accounting and Economics* 33 (2): 173–204.

Bettis, J., J. Coles, and M. Lemmon. 2000. Corporate policies restricting trading by insiders. *Journal of Financial Economics* 57 (2): 191–220.

Bradshaw, M. T., S. A. Richardson, and R. G. Sloan. 2003. Pump and dump: An empirical analysis of the relation between corporate financing activities and sell-side analyst research. Working paper, Harvard Business School.

Bradshaw, M., and R. Sloan. 2002. GAAP versus the street: An empirical assessment of two alternative definitions of earnings. *Journal of Accounting Research* 40 (1): 41–66.

Brown, K. 2002. Questioning the books: Tweaking results is hardly a sometime thing — Many firms, under diverse pressures, may play with numbers. *Wall Street Journal*, February 6, C1.

Brown, L. 2001. A temporal analysis of earnings surprises: Profits and losses. *Journal of Accounting Research* 39 (2): 221–41.

Brown, L., G. Foster, and E. Noreen. 1985. Security analyst multi-year earnings forecasts and the capital market. Studies in Accounting Research. Sarasota, FL: American Accounting Association.

Burgstahler, D., and M. Eames. 2002. Management of earnings and analysts' forecasts. Working paper, University of Washington.

Cohen, L. 1991. Low-balling: How some companies send stocks aloft. *Wall Street Journal*, May 6, C1.

Cooper, R., T. Day, and C. Lewis. 2001. Following the leader: A study of individual analysts' earnings forecasts. *Journal of Financial Economics* 61 (3): 383–416.

Cotter, J., A. I. Tuna, and P. D. Wysocki. 2004. Expectations management and beatable targets: How do analysts react to public earnings guidance? Working paper, University of Pennsylvania.

Dechow, P., A. Hutton, and R. Sloan. 2000. The relation between analysts' long-term earnings growth forecasts and stock price performance following equity offerings. *Contemporary Accounting Research* 17 (1): 1–32.

DeGeorge, F., J. Patel, and R. Zeckhauser. 1999. Earnings management to exceed thresholds. *Journal of Business* 72 (1): 1–33.

Fama, E. F., and J. D. Macbeth. 1973. Risk, return and equilibrium — Empirical tests. *Journal of Political Economy* 81 (3): 607–36.

Francis, J., and D. Philbrick. 1993. Analysts' decisions as products of a multi-task environment. *Journal of Accounting Research* 31 (2): 216–30.

Givoly, D., and C. Hayn. 2000. The changing time-series properties of earnings, cash flows and accruals: Has financial reporting become more conservative? *Journal of Accounting and Economics* 29 (3): 287–320.

Hall, B., and J. Liebman. 1998. Are CEOs really paid like bureaucrats? *Quarterly Journal of Economics* 113 (3): 653–91.

Hirshleifer, D., and S. H. Teoh. 2003. Herd behaviour and cascading in capital markets: A review and synthesis. *European Financial Management* 9 (1): 25–66.

Hutton, A. 2003. The determinants and consequences of managerial earnings guidance prior to regulation fair disclosure. Working paper, Harvard Business School.

Insider Trading Sanctions Act. 1984. Pub. L. no. 98-376, 98 Stat. 1264.

Insider Trading and Securities Fraud Enforcement Act. 1988. Pub. L. no. 100-704, 102 Stat. 4677.

Jeng, L. 1999. Corporate insiders and the window of opportunity. Working paper, Boston University School of Management.

Kasznik, R., and M. F. McNichols. 2002. Does meeting earnings expectations matter: Evidence from analyst forecast revisions and share prices. *Journal of Accounting Research* 40 (3): 727–59.

Korajczyk, R., D. Lucas, and R. McDonald. 1991. The effect of information releases on the pricing and timing of equity issues. *Review of Financial Studies* 4 (4): 685–708.

Laderman, J. L. 1998. Wall Street's spin game — Stock analysts often have hidden agenda. *Business Week*, October 1998, 148–53.

Lim, T. 2001. Rationality and analyst forecast bias. *Journal of Finance* 56 (1): 369–85.

Lin, H., and M. McNichols. 1998. Underwriting relationships and analysts' earnings forecasts and investment recommendations. *Journal of Accounting and Economics* 25 (1): 101–27.

Lucas, D., and R. McDonald. 1990. Equity issues and stock price dynamics. *Journal of Finance* 45 (4): 1019–43.

Matsumoto, D. 2002. Management's incentives to avoid negative earnings surprises. *The Accounting Review* 77 (3): 483–514.

McNichols, M., and P. O'Brien. 1997. Self-selection and analyst coverage. *Journal of Accounting Research* 35 (Supplement): 167–99.

Michaely, R., and K. Womack. 1999. Conflict of interest and the credibility of underwriter analyst forecasts. *Review of Financial Studies* 12 (4): 653–86.

Murphy, K. J. 1999. Executive compensation. In *Handbook of Labor Economics*, vol. 3, eds. O. Ashenfelter and D. Card. Amsterdam: North Holland.

Noe, C. 1999. Voluntary disclosures and insider transactions. *Journal of Accounting and Economics* 27 (3): 305–26.

O'Brien, P. 1988. Analysts' forecasts as earnings expectations. *Journal of Accounting and Economics* 10 (1): 53–83.

Richardson, S. A., S. H. Teoh, and P. D. Wysocki. 1999. Tracking analysts' forecasts over the annual earnings horizon: Are analysts' forecasts optimistic or pessimistic? Working paper, University of Michigan.

Securities and Exchange Commission (SEC). 2003. Regulation analyst certification (Regulation AC). 17 CFR part 242; release nos. 33-8193 and 34-47384. Available online at http://www.sec.gov/rules/final/33-8193.htm.

Securities and Exchange Commission (SEC). 2000. Selective disclosure and insider trading (Regulation FD). 17 CFR parts 240, 243, and 249; release nos. 33-7881, 34-43154, and IC-24599. Available online at http://www.sec.gov/rules/final/33-7881.htm.

Securities Exchange Act. 1934. C. 404, Title I, § 1, 48 Stat. 881.

Sivakumar, S., and J. Vijaykumar. 2001. Insider trading, analysts forecast revisions and earnings changes. *Journal of Accounting Auditing and Finance* 16 (2): 167–87.

Sivakumar, K., and G. Waymire. 1994. Insider trading following material news events: Evidence from earnings. *Financial Management* 23 (1): 23–32.

Skinner, D., and R. Sloan. 2002. Earnings surprises, growth expectations and stock returns or don't let an earnings torpedo sink your portfolio. *Review of Accounting Studies* 7 (2/3): 289–312.

Soffer, L., R. Thiagarajan, and B. Walther. 2000. Earnings preannouncement strategies. *Review of Accounting Studies* 5 (1): 5–26.

Teoh, S. H., I. Welch, and T. Wong. 1998a. Earnings management and the long-term market performance of initial public offerings. *Journal of Finance* 53 (6): 1935–74.

Teoh, S. H., I. Welch, and T. Wong. 1998b. Earnings management and the underperformance of seasoned equity offerings. *Journal of Financial Economics* 50 (1): 63–99.

Teoh, S. H., and T. Wong. 2002. Why new issues and high accrual firms underperform: The role of analysts' credulity. *Review of Financial Studies* 15 (3): 869–900.

The Biggest Mistakes We Teach

Forthcoming, *Journal of Financial Research* (Summer 2002)

Jay R. Ritter
Cordell Professor of Finance
University of Florida
ritter@dale.cba.ufl.edu
http://bear.cba.ufl.edu/ritter

December 3, 2001

Presented at the Southern Finance Association Annual Meetings, November 30, 2001. The comments of Joel Houston, Michael Ryngaert, Richard Warr, and especially Jeremy Siegel are appreciated.

I. Introduction

When I started to teach at the University of Pennsylvania's Wharton School over twenty years ago, I used the very first edition of the Brealey and Myers' textbook. The book had some mistakes in it, as almost all books do. For example, the first two editions had an incorrect formula for the valuation of warrants. I taught the incorrect formula for several years before a perceptive student asked a question that exposed the mistake. But I don't want to dwell on technical errors. Instead, I want to focus on some of the conceptual mistakes that dominate the received body of wisdom in the academic finance profession.

II. The Relative Risk of Stocks and Bonds

Almost all finance textbooks prominently feature the historical returns provided by Ibbotson Associates. These numbers show that since 1926, stocks have produced higher average annual returns than bonds, and that stocks are riskier than bonds. This is consistent with equilibrium risk-return models. There are three problems with this evidence that stocks are riskier than bonds, however.

First, the use of annual holding periods. There is no theoretical reason why one year is the appropriate holding period. People are used to thinking of interest rates as a rate per year, so reporting annualized numbers makes it easy for people to focus on the numbers. But I can think of no reason other than convenience for the use of annual returns. If returns follow a random walk, then whether a one year holding period is used, or a shorter or longer period is used, makes no difference. But if there is mean reversion or mean aversion in the data, then the risk of one class of securities relative to another depends on the holding period.

Second, the use of arithmetic, rather than geometric returns. The relation between the arithmetic (simple) average and the geometric (compounded) average is given by the formula

$$r_{arith} = r_{geo} + 1/2\sigma^2$$

The higher is the variance rate, the larger will be the difference between the arithmetic and geometric returns. For stocks, the difference between the arithmetic and geometric averages is about 2% per year. For bonds, the difference is much smaller. As a result, the performance of stocks relative to bonds looks better when arithmetic averages are compared than when geometric averages are compared. Now, if stock and bond returns follow a random walk, the use of annual arithmetic returns is appropriate. But if there is mean reversion or mean aversion, then the use of arithmetic returns over longer time periods is not appropriate. With mean reversion, the multi-period arithmetic return will be closer to the geometric return.

Third, the use of nominal, rather than real returns. People are concerned about the consumption bundle that they can consume. The only reason that nominal returns, rather than real returns, should be reported in textbooks is simplicity. But this simplicity comes at a cost. If stocks are good short-term hedges against inflation, they could have a higher variance of nominal returns and yet offer a lower variance of real returns. In fact, stocks are bad short-term hedges against inflation. On theoretical grounds, it is the standard deviation of real returns that is relevant.

Figure 1 provides an updated version of Figure 2-4 in Jeremy Siegel's *Stocks for the Long Run*, showing the standard deviation of real returns for different holding periods, using data starting in 1802. For a one-year holding period, stocks are twice as risky as bonds. For holding periods of twenty or more years, however, stocks are less risky than bonds.

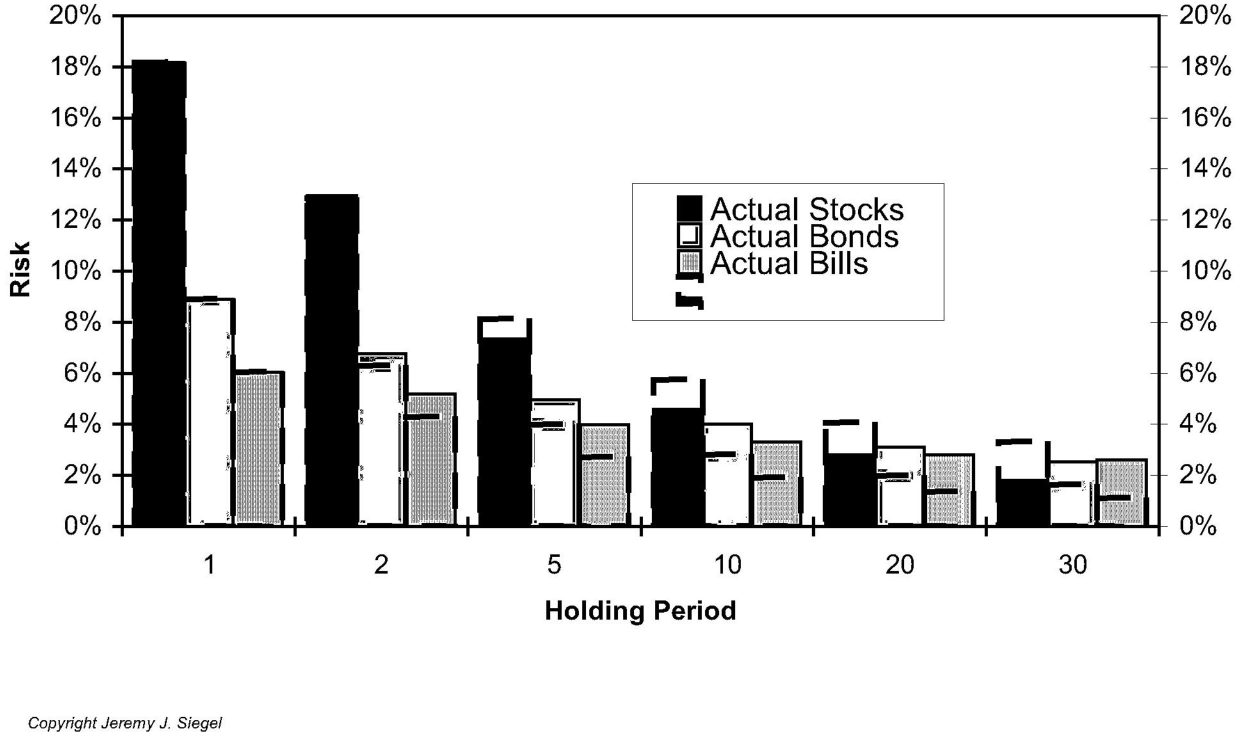


Figure 1: The annualized standard deviation of compounded real holding-period returns from Janaury 1802 to September 2001. For example, a two-year buy-and-hold real return of 21% would have an annualized compounded real return of 10%. For the sample period, there are 199 overlapping two-year returns, from which 199 annualized numbers are calculated. The bars represent these actual standard deviations. The dashed bars represent what the standard deviations would be if the one-year standard deviations are divided by the square root of the holding period, which is the random walk assumption. This is an updated version of Figure 2-4 from Siegel (1998), supplied by Jeremy Siegel.

Why is this so? Well, although stocks are a bad hedge against inflation in the short-run, they are a good hedge against inflation over a longer period of time, such as five years. This pattern is a major contributor to the negative autocorrelation of real stock returns that exists over a five-year horizon. In other words, real stock returns show a tendency towards mean-reversion. This makes stocks less risky over a T-year holding period than would be suggested by multiplying the annual variance by T. If there is no mean reversion, the T-period variance of returns, σ^2_T, is equal to T times the variance of single-period returns, σ^2. If one uses monthly returns data, however, researchers generally find that $\sigma^2_T < T\sigma^2$ when using a market index when T is greater than 24 months.

I can think of another reason why real stock returns are negatively autocorrelated at three-to-five year horizons. If individuals put too much weight on recent evidence, then they will put more money into stocks after stocks have done well, pushing up the prices even further. Similarly, after stocks have done poorly, they will pull money out of stocks, depressing prices further. This is an example of the representativeness heuristic. People put too much weight on recent evidence. This is also known as the fallacy of small numbers.

In contrast to stocks, the real returns on nominal bonds show no tendency towards mean reversion. In fact, there is a slight tendency towards mean-aversion, making them more risky the longer the holding period. But the big risk with nominal bonds comes from a hyper-inflation. Fortunately, the U.S. has never had a hyper-inflation, but other countries have. In a hyper-inflation, stocks typically have negative real returns, but then recover, at least partially. Bonds get wiped out in real terms, and once this occurs, you can never recover.

Stocks are riskier than bonds for short holding periods. But it is not at all obvious that this is true for long holding periods, either historically or in the future.

III. Estimating the Future Equity Risk Premium

The equity risk premium is the difference in returns between stocks and safe assets, such as Treasury bills. There are three approaches to estimating the equity risk premium on a point-forward basis. The first approach is to extrapolate historical returns. The second approach is to use a theoretical model of what the equity premium should be, given plausible assumptions about risk aversion. The third approach is to use forward-looking information such as the current dividend yield and interest rates.

Many textbooks encourage students to use the historical arithmetic equity risk premium of 9% for computing the cost of equity capital. Ivo Welch's recent survey of financial economists indicates that most finance professors extrapolate the historical average, too, although many shade it down to about 7%, perhaps due to concerns about survivorship bias. The numbers that I am about to compute using forward-looking information suggest that 1% is a more defensible number.

Before doing so, let me point out how extrapolating historical numbers can result in numbers that are nonsensical. If one were estimating the equity risk premium for Japan at the end of 1989, using the historical data starting when the Japanese stock market reopened after World War II, one would produce an equity risk premium of more than 10%. But at the end of 1989, the Japanese economy was booming, corporate profits were high, and the market's price-earnings ratio was over 60. At the time, it was the conventional wisdom that the cost of equity capital for Japanese corporations was low. It *cannot* be the case that the cost of equity capital is low *and* the equity risk premium is high. But it *can* be the case that the historical equity premium is high, and the expected equity risk premium for the future is low.

If a theoretical model is used for what the equity risk premium should be, one comes up with a number in the vicinity of 2% if geometric returns are used, or 4% if arithmetic returns are used. This is the approach used by Mehra and Prescott (1985) in their famous paper.

The first forward-looking approach to estimate the future real return on equities is to look at the market's earnings yield. The earnings yield is just the reciprocal of the P/E ratio. Now, one must normalize earnings because earnings may be temporarily high or low due to business cycle effects. Historically, the earnings yield has averaged 7%. Not coincidentally, the average compounded real return on equities has averaged 7%. This historical average of 7% is composed of a dividend yield of 4.5% and a real capital gain of 2.5%.

Today, the earnings yield is in the vicinity of 4%, once one smoothes out business cycle effects. This generates a real return on equities, on a point-forward basis, of about 4%, which is below the historical average. The lower forecast today is because the P/E ratio is higher than the historical average of about 14. The higher P/E ratio today also results in a lower dividend yield. Today, the dividend yield is about 1.5%. The dividend yield is low both because the P/E ratio is high, and the payout ratio of dividends to earnings is relatively low. The dividend payout ratio is low partly because of the increase in share repurchases. Because of share repurchases, expected real capital gains have increased. But employee stock options have also become more popular, and this dilution partly offsets the effect of share repurchases. A 2.5% real capital gain per share plus a 1.5% dividend yield produces a 4% per year real return on equities.

The second forward-looking approach is to use the Gordon dividend growth model. Using this model, which is a rearrangement of the growing perpetuity formula $P_0 = Div_1/(r - g)$, one gets that

$$r = \text{the dividend yield} + g$$

where g is the growth rate of dividends per share. If the dividend yield stays constant over time, then the growth rate of dividends per share will be the same as the growth rate of the stock price.

What is a plausible estimate of g? If aggregate dividends grow at 2.5%, and the aggregate dividend/labor income ratio for the economy stays constant, this would imply that real labor income grows at 2.5%. If the population grows at 1%, this would imply that per capita income grows at 1.5% per year. This is equal to the historical average long-term growth rate of about 1.5% in developed countries, according to Prichett (1997). A 1.5% per year growth rate means that real per capita income will double every 47 years. If the net effect of share repurchases and option dilution adds 1% to per share growth, then a growth rate of real dividends per share of 2.5% can be justified. Adding a 1.5% dividend yield to this gives a 4% real return on equities in the future.

Since 1997, the U.S. Treasury has issued inflation-indexed bonds, commonly known as TIPS, for Treasury Inflation-Protected Securities. These bonds do offer protection against inflation risk. Many textbooks do not even acknowledge the existence of this important asset class.

The Ibbotson numbers show that the historical real return on bonds has been about 1%. But today, TIPs are yielding real returns of about 3.3%. If the expected real return on equities is 4% and the real return on inflation-indexed bonds is 3.3%, the equity risk premium is only 0.7%. In round numbers, 1%. The equity premium has gotten squeezed from the top (low future real returns on stocks) and the bottom (a higher real return on bonds).

I think that textbooks should present historical returns, but should focus on the Gordon dividend growth model for estimating the future equity risk premium. For predicting future dividend growth rates, all one has to do is assume an economy-wide growth rate and then assume that the ratio of labor income to capital income is a constant. Fama and French (2002) and Jagannathan, McGratton, and Scherbina (2000), among others, also adopt the Gordon dividend growth model framework and conclude that the equity risk premium is now in the vicinity of 1%, far below the historical average.

IV. The Fed Model

The so-called Fed Model states that the stock market is fairly valued when the earnings yield on stocks is equal to the interest rate on bonds. This model for valuing stocks is based on the empirical regularity that is illustrated in Figure 2.

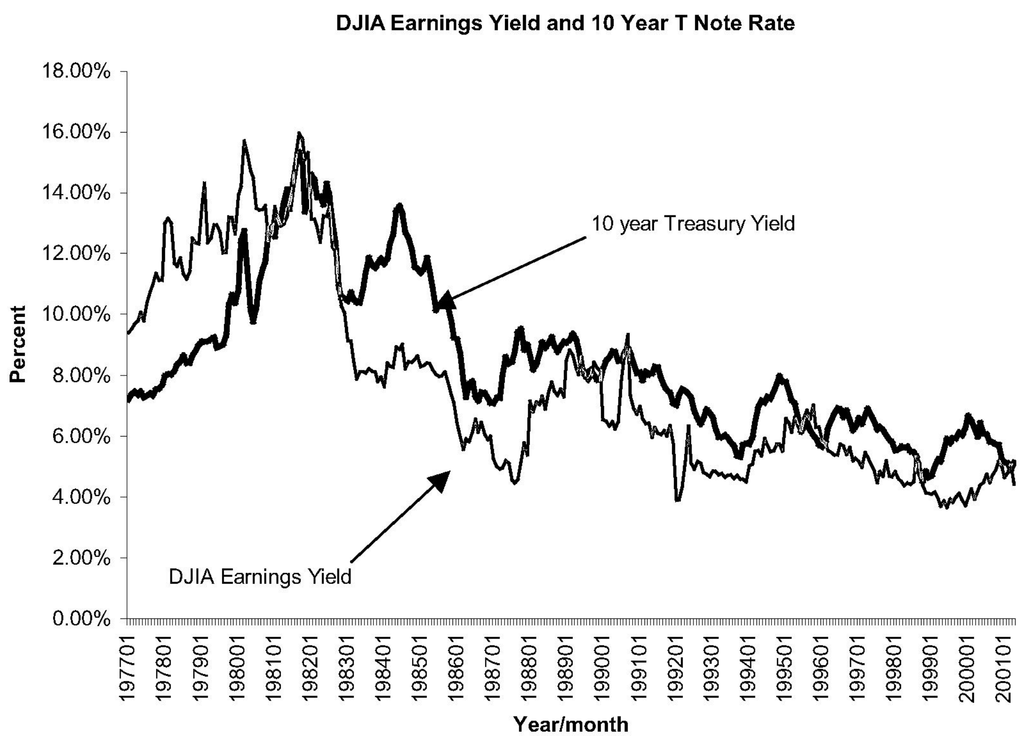


Figure 2: Monthly values of the earnings yield (last fiscal year's earnings) on the Dow Jones Industrial Average and the nominal yield on 10-year Treasury securities.

Empirically, this is a model that works very well. But on theoretical grounds, if most of the variation in nominal interest rates comes from changes in expected inflation rather than changes in real rates, the model should not work well. In fact, the strong positive correlation should theoretically be negative, in an efficient market. The logic was first pointed out by Modigliani and Cohn in their 1979 *FAJ* article, and is reiterated in my paper with Richard Warr in the March 2002 *JFQA*. The logic is that, for firms with debt in their capital structure, earnings are depressed by high nominal interest payments. The part of the nominal interest payment that goes to compensate bondholders for inflation reflects the decline in the real value of the liabilities of the firm. Accountants measure the cost to equityholders from the interest payments, but they don't measure the benefit to equityholders from the decline in the value of the firm's real liabilities. Thus, in an inflationary environment, accounting earnings underestimate the true economic earnings of a firm. Since accounting earnings are used to calculate the price-earning (P/E) ratio, the more economic earnings are understated, the higher should be the P/E ratio.

Now, inflation distorts accounting earnings in other ways, and the tax system is not inflation-neutral. But when Richard Warr and I adjust for these other effects, we conclude that the net impact is that P/E ratios should be higher, not lower, in periods of high inflation. This is exactly the opposite of the empirical evidence.

I think that there is a complacency in the profession. If we have an empirical pattern that is difficult to reconcile with theory, we shy away from saying that the market gets it wrong. Instead, we search for other explanations or just ignore the inconvenient facts.

The Fed model is typically not discussed in textbooks. But it is frequently discussed in the financial press, and there is never any discussion of why the empirical relation is inconsistent with rational valuation. Adjusted for business cycle effects, the earnings yield on stocks is an estimate of the expected real return on stocks.[1] The earnings yield is *not* an estimate of the expected *nominal* return on stocks. For the earnings yield to move one-for-one with the nominal bond yield, as the Fed model would have it, one has to assume that the nominal yield on bonds equals the real return on stocks. This is why the empirical success of the Fed model is inconsistent with rational valuation.

V. The Limits to Arbitrage and Market Efficiency

Securities markets in the United States are very good at getting the little things right. It is incredibly difficult to find high-frequency arbitrage opportunities that persist. But in my opinion, the profession has made a serious error in jumping to the conclusion that if the market gets the little things right, it must get the big things right. Low-frequency events are not amenable to formal statistical tests. By definition, they don't repeat themselves frequently. What makes it difficult to separate out overreactions that slowly correct themselves from rational time-variation in equilibrium expected returns is that the market gets overvalued when there are legitimate grounds for optimism, and undervalued when there are legitimate grounds for pessimism.

[1] Note that every textbook points out that the earnings yield on a stock is not the cost of equity capital for the firm, because earnings growth rates for firms vary all over the map. But the economy's growth rate of earnings does not vary much over time, once one accounts for business cycle effects. So the "normalized" earnings yield on the market is a good estimate of the cost of equity capital, in real terms, for the market as a whole.

By low-frequency events, I am referring to things like the October 1987 stock market crash, the Japanese bubble of the 1980s, and the TMT (technology, media, and telecom) bubble of the late 1990s.

Market efficiency does not just mean the lack of arbitrage profits. Just because it is difficult to design and implement strategies that will reliably make positive risk-adjusted profits does not mean that large misvaluations are not common. As Shleifer and Vishny (1997) have pointed out, taking positions in misvalued securities is extremely risky. For instance, if one shorted overvalued Japanese stocks at the beginning of 1988, one would have lost substantial money over the next two years. An investor who did this might not have had any capital left when the bubble finally burst starting in January of 1990.

Similarly, money managers that bet against overvalued internet stocks in early 1999 suffered huge losses before the TMT bubble burst starting in March 2000. Few of these investors had any capital left in March 2000. As with the Japanese bubble, unless one had the foresight to avoid taking a position when the misvaluations were large, and wait until the misvaluations became very large, you would have been wiped out. Being right in the long run is no consolation if you have lost everything in the short run.

But I am hard-pressed to find a discussion along these lines in most textbooks. Instead, the evidence on high-frequency efficiency is typically fallaciously applied to assert that low-frequency inefficiencies won't exist.

VI. Dividend Policy

The chapter on dividend policy should be called payout policy. There are two distinct issues-- the form of payout, and the level of payout. In the days of M&M, these were pretty much one and the same. But since 1984, they have been very different. The typical textbook covers the Modigliani and Miller theorem, taxes, and signaling, and then at the end of the chapter adds a few paragraphs on share repurchases. Instead, I would suggest that the first half of the chapter should be devoted to what determines the level of cash payouts, and the second half should be devoted to the choice between share repurchases and dividends. The empirical evidence is that taxes are at best a second-order consideration in determining the form of payout. In particular, any tax-based model would predict that there should have been much more share repurchases prior to the 1986 tax reform act, because capital gains had been given preferential tax status. Shefrin and Statman's 1984 *Journal of Financial Economics* article giving behavioral reasons for cash dividends is barely mentioned, if it is mentioned at all, in most textbooks.

I suspect that if most of us were writing a textbook from scratch today, the chapter on payout policy would look very different than the one that appears in textbooks. There is a strong path-dependency involved. Even if a textbook author wants to make a major change, most professors don't want to have to revise their lecture notes.

VII. Lease Finance

Most textbooks cover leasing before they cover options. Many leases give the lessee the right to buy the item that they have leased at the end of the lease, at a fixed exercise price. This option is valuable. But most textbooks ignore it, because they haven't covered option pricing theory yet.

Similarly, most textbooks cover issuing equity before options are covered. Many of these textbooks cover rights offerings in their chapter on issuing equity or raising capital. But because they haven't covered options yet, they don't note that a right is just a warrant. So they don't give the correct formula for valuing a right that is not deep in the money.

The deferral of the options chapter until late in the book has other costs. In one prominent textbook (I won't mention names, to protect the guilty), convertible bonds are covered before option pricing is covered. The gyrations that the textbook has to go through are funny, except that students don't get the humor.

VIII. Conclusions

I've taken issue with the way we as a profession teach certain things, and the way that textbooks present them. These are some of my pet peeves. I'm sure that each of us could make up a list. But I have to concede that I find it a lot easier to criticize others than to do it right myself. I have no intention of writing a textbook. And even if I did, and got a lot of things right that other textbooks get wrong, I'm sure that I would introduce different mistakes.

About seven years ago I attended an NBER meeting where Michael Jensen was one of the speakers. Jensen received his Ph.D. from Chicago in 1968. I received my Ph.D. from Chicago in 1981, and by that time a number of Jensen's articles were on the reading lists. At the NBER meeting, Jensen said that he had come to realize that most of what he learned in graduate school was wrong. Well, I feel that way, too. Twenty years from now, I expect that my former doctoral students will be saying that a lot of what they learned in graduate school was wrong. I just wish that I knew now which things that I'm teaching are wrong, rather than having to wait twenty years to find out.

References

Fama, Eugene F., and Kenneth French, 2002, The Equity Premium, *Journal of Finance* 57, forthcoming.

Jagannathan, R., E. R. McGrattan, and A. Scherbina, 2000, The Declining U.S. Equity Premium, *Federal Reserve Bank of Minneapolis Quarterly Review* 24, 3-19.

Mehra, Raj, and Edward Prescott, 1985, The Equity Risk Premium: A Puzzle, *Journal of Monetary Economics* 15, 145-161.

Modigliani, F., and R. A. Cohn, 1979, Inflation, Rational Valuation and the Market, *Financial Analysts Journal* 35, 24-44.

Prichett, Lant, 1997, Divergence, Big Time, *Journal of Economic Perspectives* 11, 3-17.

Ritter, J. R., and R. S. Warr, 2002, The Decline of Inflation and the Bull Market of 1982-1999, *Journal of Financial and Quantitative Analysis* 37, forthcoming.

Shefrin, H. M., and M. Statman, 1984, Explaining Investor Preference for Cash Dividends, *Journal of Financial Economics* 13, 253-282.

Shleifer, A., and R. Vishny, 1997, The Limits of Arbitrage, *Journal of Finance* 52, 35-55.

Siegel, J. J., 1998, *Stocks for the Long Run, Second Edition.* New York: McGraw-Hill.

Welch, Ivo, 2000, Views of Financial Economists on the Equity Premium and on Professional Controversies, *Journal of Business* 73, 501-537.

VOLUME 24 | NUMBER 3 | SUMMER 2012

Journal of APPLIED CORPORATE FINANCE

A MORGAN STANLEY PUBLICATION

In This Issue: **Growth, Value, and Payout Policy**

Morgan Stanley

Is Economic Growth Good for Investors?[1]

by Jay R. Ritter, University of Florida

Economic growth is widely believed to be good for stock returns, and forecasts of growth are a staple of international asset allocation decisions. Investing in emerging markets with good long-term growth prospects, such as China, is widely viewed as more attractive than investing in countries like Argentina, with prolonged periods of low growth that are expected to persist. But does economic growth always—or even generally—benefit stockholders?

Surprisingly, the answer is no, on both theoretical and empirical grounds. For 19 countries with continuously operating stock markets during the 112-year stretch from 1900 through the end of 2011, the cross-sectional correlation between returns and the growth rate of per capita gross domestic product (GDP) is negative—to be more precise, the correlation coefficient is -0.39. This negative correlation suggests that investors in 1900 would actually have been better off investing in the companies of nations that ended up experiencing lower per capita growth than in those countries that enjoyed higher average growth rates.

In 1900, most of these 19 countries were considered economically advanced, or "developed," nations, and their publicly traded companies likely accounted for 90% or more of the market value of the world's equities at that time. This negative correlation between per capita GDP growth and equity returns has been experienced not only by developed countries, however, but by developing economies as well. For 15 emerging markets during the 24-year period from 1988 to 2011—including the BRIC countries of Brazil, Russia, India, and China—the correlation is a remarkably similar -0.41.

I am not arguing that economic growth is bad. There is ample evidence that people who live in countries with higher incomes have higher standards of living and longer life spans. But even though consumers and workers typically benefit from economic growth, the owners of capital do not necessarily benefit. As I will discuss later, countries can grow rapidly—and for considerable periods of time—by applying more labor and capital, but without the owners of capital earning high returns. And it's much the same story with economic growth due to technological advances: Unless technological change comes from companies with some kind of pricing power, the resulting improvements in productivity typically end up contributing mainly to higher standards of living for workers and consumers, and not to higher shareholder returns.

In this article, I start by documenting the negative correlations between long-run economic growth and stock returns for both developed countries and emerging markets—and go on to offer a number of explanations for this somewhat surprising relationship. Then I explain why the standard of living in a country can grow rapidly without investors earning high—or, in many cases, even just competitive—returns. In the final section, after relating past per capita income growth to historical stock returns, I consider the relation between economic growth and *future* expected returns. The major determinant of future returns is the earnings yield—the reciprocal of the price-earnings ratio—that investors are paying today. From a managerial perspective, I focus particular attention on two variables: the percentage of earnings that companies reinvest in the business, and the rate of return on such reinvestment. As finance professors have long taught their students, the key to adding value for shareholders is for companies to invest in all positive net present value (NPV) projects—while at the same time committing to return to their investors through dividends and stock buybacks all capital and cash flow that cannot be so reinvested.

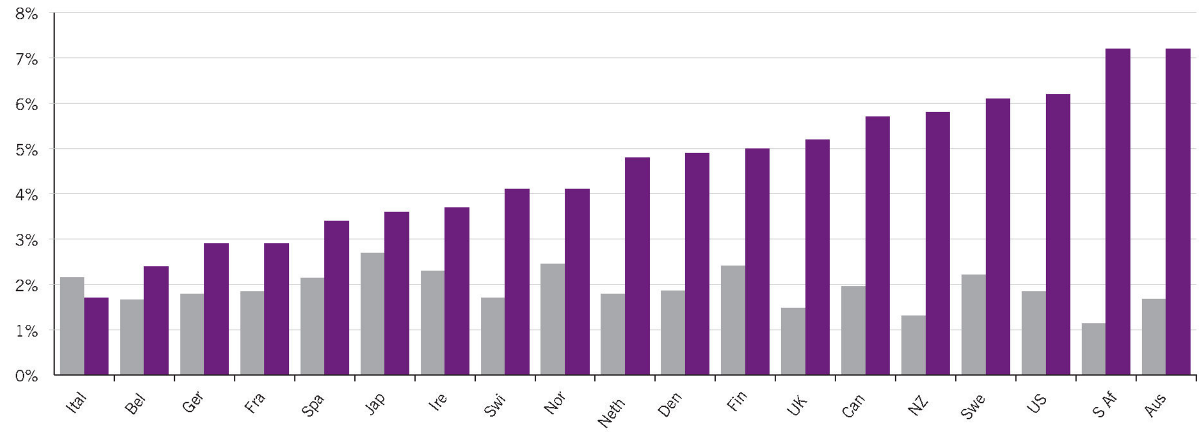
The Negative Correlation Between GDP Growth and Real Stock Returns

To the best of my knowledge, the first mention of a negative cross-country correlation between real per capita GDP growth and real stock returns was by Jeremy Siegel in the second edition (2002) of his book, *Stocks for the Long Run*. Siegel's claim, however, was based on data that go back only as far as 1970.

In Table 1 and Figure 1, I summarize the existing evidence showing the negative correlation from 1900-2011

1. This paper updates and extends through 2011 the findings that were first presented in my 2005 *Pacific-Basin Finance Journal* article, "Economic Growth and Equity Returns," which contains a more complete list of citations and references. I want to thank Leming Lin for excellent research assistance, and the editor, Don Chew, for extensive suggestions and guidance. Comments from Jeremy Siegel and participants at the TAPMI International Conference in Banking and Finance in Bangalore and seminars at the Australian National University and the University of Melbourne are also appreciated.

Figure 1



Real per capita GDP growth rate per annum (on left in gray) and real equity return per annum (on right, in purple), 1900-2011. The real return data (dividends plus capital gains, adjusted for inflation, in local currency units) are from Dimson, Marsh and Staunton (2012). Real per capita GDP growth rates are from the World Bank, Dimson, Marsh and Stanton (2012), and Maddison (2010).

between real per capita gross domestic product (GDP) growth and real stock returns for 19 mostly developed countries with continuously operating stock markets since 1900.[2] Throughout this article, all stock returns and income growth rates are expressed in dollars of constant purchasing power—in other words, they have been adjusted for inflation. The source of the average long-run stock returns is the *Credit Suisse Global Investment Returns Sourcebook 2012*, which contains the most recent annual update by Elroy Dimson, Paul Marsh, and Mike Staunton of London Business School of findings that were first published in their 2002 book *Triumph of the Optimists*.[3] The book reported finding a negative correlation between real stock returns and real per capita economic growth for 16 countries during the period 1900-2001.[4] And since the publication of their book, Dimson, Marsh, and Staunton have presented extensive additional analysis of the negative correlation for additional countries and other time periods in their 2005 and 2010 *Yearbooks*.

As stated earlier, the correlation of real per capita GDP growth and real stock returns for 19 countries with stock markets since 1900 is -0.39 (p-value=0.10) when the returns are measured in local currencies. When the returns are adjusted for changes in the exchange rate relative to the U.S. dollar, so that they represent what a U.S. investor would have received, Table 1 reports that the correlation changes slightly, to -0.32 (p-value=0.14). The import of these findings is that an investor would have been better off avoiding countries where per capita GDP rose the most and investing in countries with slower per capita growth.

As can also be seen in the table, long-run average per capita real GDP growth rates range from a low of 1.1% for South Africa to a high of 2.7% for Japan. The average compounded real returns on equities stretch on the low end from 1.7% for Italy to over 7% for Australia and South Africa, with returns in the U.S., Canada, and the U.K. all being in the 5% to 6% range.

What do the high-return countries have in common? First of all, the top seven countries—Australia, South Africa, the United States, Sweden, New Zealand, Canada, and the United Kingdom—all have had the good fortune to avoid having major wars fought on their own soil in the last century, a misfortune that befell most of the continental European countries. Second, the high-return countries, with the exception of Sweden, are English-speaking with traditions of English common law and, apart from South Africa, long histories of democratic government and universal suffrage. Third, and also worth noting, several of these countries have had economies where the natural resources sector has played an important part in their success.

Alongside the long-run average stock returns and per capita growth rates, Table 1 also reports the average dividend yield and growth rate of real dividends per share for the same 19 countries. One of the most notable patterns is the strong association between high dividend growth rates and high overall stock returns. In one sense, such an association is not surprising in that growing dividends tend to reflect increases

2. Some of the markets have temporarily suspended trading due to war, etc. For example, the U.S. stock market closed on Sept. 11-14, 2001 following terrorist attacks. None of the 19 markets saw investors wiped out, however, unlike Russia in 1917 and China in 1949. Note that in the case of both Russia and China, both bond investors and equity investors were expropriated.

3. Also, see Elroy Dimson, Paul Marsh, and Mike Staunton, "Global Evidence on the Equity Risk Premium," *Journal of Applied Corporate Finance* (Fall 2003).

4. For Germany, Dimson, Marsh, and Staunton exclude the 1922-1923 hyperinflation years.

Table 1 **Real Annual Per Capita GDP Growth Rates and Stock Returns, 1900-2011**

Country	Real per capita GDP growth	Mean geometric real return: Local currency	Mean geometric real return: U.S. dollars	Real dividend per share growth	Dividend yield
Australia	1.68%	7.2%	7.3%	0.99%	5.7%
South Africa	1.13%	7.2%	6.4%	1.05%	5.8%
United States	1.85%	6.2%	6.2%	1.31%	4.2%
Sweden	2.21%	6.1%	6.2%	1.80%	4.0%
New Zealand	1.30%	5.8%	5.5%	1.17%	5.4%
Canada	1.96%	5.7%	5.7%	0.67%	4.4%
United Kingdom	1.48%	5.2%	5.2%	0.45%	4.6%
Finland	2.41%	5.0%	5.1%	0.23%	4.8%
Denmark	1.86%	4.9%	5.4%	-0.96%	4.6%
Netherlands	1.78%	4.8%	5.2%	-0.61%	4.9%
Switzerland	1.70%	4.1%	5.1%	0.47%	3.5%
Norway	2.45%	4.1%	4.4%	-0.07%	4.0%
Ireland	2.30%	3.7%	4.0%	-1.29%	4.5%
Japan	2.69%	3.6%	4.2%	-2.36%	5.2%
Spain	2.14%	3.4%	3.5%	-0.58%	4.2%
France	1.85%	2.9%	2.8%	-0.75%	3.8%
Germany	1.78%	2.9%	3.2%	-1.27%	3.7%
Belgium	1.66%	2.4%	3.0%	-1.48%	3.7%
Italy	2.15%	1.7%	1.8%	-2.21%	4.0%
Correlation of growth and returns		-0.39	-0.32		
p-value		(0.10)	(0.18)		

For real per capita GDP growth per year, data come from an updated version of Angus Maddison (1995) Monitoring the World Economy 1820-1992 Paris: OECD Development Centre Studies, as explained in Appendix Table A-1 for 1900-2008, and from the World Bank's World Development Indicators for 2008-2011. Real per capita income is expressed in terms of dollars of 1990 Geary-Khamis dollars (purchasing power parity-adjusted) through 2008 multiplied by the ratio of 2011/2008 real per capita income in local currency units from World Development Indicators to obtain the 2011 number, and converted into an annualized number. The South African GDP numbers start in 1913 rather than 1900. The geometric mean annual real dividend growth rates, dividend yields, and real returns (dividends plus capital gains) per year from Dimson, Marsh, and Staunton (2012) are used for 19 countries for the 112 years from 1900-2011. The equally weighted mean real return is 4.6% per year in local currency units and 4.7% per year in U.S. dollars, and the mean per capita growth rate of real GDP is 1.8% per year.

in earnings per share. But there is likely to be another effect at work here—namely, the role of dividends (and, in the case of the U.S., stock repurchases) in limiting what might be called the "overinvestment problem," or the pursuit of growth-for-growth's sake.

Take the case of Japan, where average growth in dividends per share has actually been negative in real terms (-2.4% per year) at the same time the country was achieving the highest rate of growth (2.7%) of per capita GDP of any of the countries. Japanese policymakers have long professed their commitment to growth and full employment—when necessary, at the expense of corporate profitability—and this commitment is reflected in the negative dividend growth and, until 1994, a ban on corporate repurchases of stock. In this sense, Japanese companies' reluctance to pay out corporate cash reflects what has amounted to a national policy goal of using corporate assets to preserve growth and employment. But, as policymakers have begun to recognize, the shareholder losses resulting from this pursuit of growth *at all cost* have arguably played a major role in the country's relatively poor economic performance since 1990.

But what happens if we focus on a shorter, and more recent, time period? Table 2 reports the mean geometric real returns and growth rates of real per capita GDP for the period 1970-2011, with Austria and Singapore added to the 19 countries used in Table 1. Over the 42 years since 1970, the correlation between per capita GDP growth and real stock returns has been essentially zero for these countries, whether returns are measured in local currencies or U.S. dollars.[5]

The findings summarized thus far apply to mainly developed economies. What about developing economies?

Table 3 reports, and Figure 2 shows, for the more recent period 1988-2011, the mean geometric real return and the mean growth rate of real per capita GDP for 15 countries

5. Dimson, Marsh, and Staunton (2005, Chapter 3, Chart 31) also show that, for some combinations of countries and time periods, the correlation of real per capita GDP growth and real equity returns has been zero or even positive. Dimson, Marsh, and Staunton (2010) also report that the negative correlation between stock returns and economic growth becomes positive if we use *total* GDP growth instead of per capita GDP growth. As a matter of arithmetic, this change reflects the tendency of some countries with high stock returns to have high population growth rates. Most notable is South Africa, which has a higher birth rate than that of any of the other countries listed, as well as substantial immigration from neighboring African countries with a lower standard of living (and, in the case of Mozambique, prolonged civil wars). Because people tend to move *from* poor countries *to* richer countries, and people in richer countries tend to have lower birth rates, the population growth rates are causally related to the level of real incomes at the end of the sample period. (Table A-1 in the Appendix reports the cumulative and per annum population growth over 1900-2011 for the 19 countries used in Table 1.)

Table 2 **Real Per Capita GDP Growth Rates and Stock Returns for 21 Countries, 1970-2011**

Country	Mean geometric real per capita GDP growth	Mean geometric real return	
		Local currency	U.S. dollars
Australia	1.8%	3.6%	4.7%
Austria	2.3%	2.3%	3.5%
Belgium	2.0%	5.4%	6.2%
Canada	1.7%	5.3%	5.4%
Denmark	1.5%	6.8%	8.0%
Finland	2.4%	7.9%	8.5%
France	1.8%	4.6%	5.1%
Germany	1.7%	5.8%	4.9%
Ireland	3.3%	3.1%	4.2%
Italy	1.8%	0.3%	0.7%
Japan	2.0%	2.3%	4.6%
Netherlands	1.9%	6.2%	7.2%
New Zealand	1.2%	4.1%	4.9%
Norway	2.4%	5.6%	6.7%
Singapore	5.1%	5.9%	6.6%
South Africa	0.6%	6.9%	6.3%
Spain	2.0%	2.9%	4.5%
Sweden	1.8%	8.8%	8.8%
Switzerland	1.0%	4.6%	6.7%
United Kingdom	2.0%	4.9%	5.6%
United States	1.7%	4.9%	4.9%
Correlation of real growth and real returns		-0.04	0.01
p-value		(0.87)	(0.95)

Geometric mean real annual GDP per capita growth rates (using constant local currency units) for the 42 years from 1970-2011 come from the *World Bank's World Development Indicators* (WDI). Geometric mean real annual stock returns come from Datastream, where the MSCI total return indices with dividends reinvested are used. Inflation adjustments for stock returns are made using December to December changes in the CPI. The mean real return is 4.9% per year in local currencies and 5.6% per year in U.S. dollars and the mean real per capita GDP growth rate is 2.0% per year.

that, 24 years ago in 1988, were generally viewed as emerging markets. The group includes the four BRIC countries, even though for these cases the MSCI stock return series start later than 1988. In fact, China and Russia did not even have stock markets in 1988; and almost no one predicted the fall of the Berlin Wall a year later and the collapse of the Soviet Union. For these 15 countries, the correlation is -0.41 (p=0.13) in local currency units and -0.47 (p=0.08) in U.S. dollars.

In China, the combination of high economic growth (over 9% on average) with low stock returns (-5.5% per year) is especially notable, particularly considering the fact that China's stock market grew from almost nothing in 1993 to a market value of approximately $3 trillion at the end of 2011. Much of the growth in China's aggregate market cap is attributable to the expansion of the number of listed companies, resulting in part from several thousand initial public offerings (IPOs), including those of China's four largest state-owned banks.

Economic Growth and Stock Returns

What might explain this negative correlation between real stock returns and real per capita GDP growth?

One possibility is that part of the negative correlation reflects the tendency of investors to build expectations for high growth into prices at the start of the period. This is a major reason why the returns on Chinese stocks during the period 1993-2011 were so low (again, -5.5%). At various times since 1993, the price-earnings (P/E) multiples of Chinese company stocks reached extraordinary levels, followed by earnings disappointments and low reported returns on capital. When one uses 112 years of data, however, the effects of such anticipation on average realized returns should be fairly modest. For example, even if the stock price multiples at the beginning (1900) were twice as high in one country as another, the compounded average annual return would have been reduced by only about 0.6% per year.[6]

But that said, I think there is a general tendency for

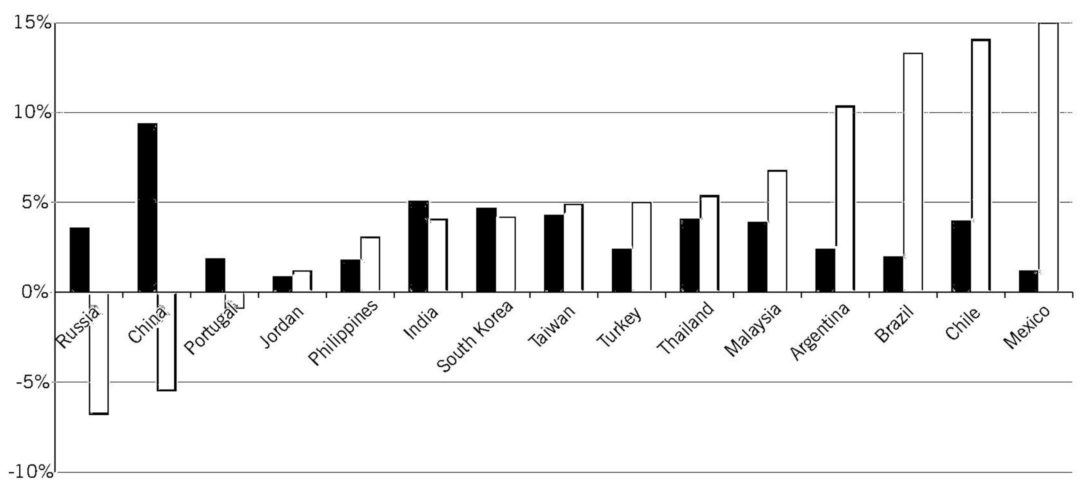
6. $1.006^{112} = 1.954$, or approximately 2. If country A and country B both give stock market investors terminal inflation-adjusted wealth of 100 (capital gains plus reinvested dividends) at the end of 2011, but in 1900 country A required an investment of 2 and country B required an investment of 1, the compounded annual returns are, respectively, 3.6% per year for country A and 4.2% per year for country B, a difference of 0.6% per year.

Table 3 **Real Stock Returns and Per Capita GDP Growth for 15 Countries for (up to) 24 years**

Country	Years	Mean geometric real per capita GDP growth	Mean geometric real return: Local currency	Mean geometric real return: U.S. dollars
Argentina	1988-2011	2.4%	10.4%	12.9%
Brazil	1993-2011	2.0%	13.3%	10.7%
Chile	1988-2011	4.0%	14.1%	15.2%
China	1993-2011	9.4%	-5.5%	-5.7%
India	1993-2011	5.1%	4.1%	4.1%
Jordan	1988-2011	0.9%	1.2%	0.3%
Malaysia	1988-2011	3.9%	6.8%	5.9%
Mexico	1988-2011	1.2%	15.0%	17.1%
Philippines	1988-2011	1.8%	3.1%	4.3%
Portugal	1988-2011	1.9%	-0.9%	0.0%
Russia	1995-2011	3.6%	-6.8%	-2.2%
South Korea	1988-2011	4.7%	4.2%	4.1%
Taiwan	1988-2011	4.3%	4.9%	2.8%
Thailand	1988-2011	4.1%	5.4%	5.2%
Turkey	1988-2011	2.4%	5.0%	6.9%
Correlation of real growth and real returns			-0.41	-0.47
p-value			(0.13)	(0.08)

For real per capita income, the average level of the price level in a year is used to convert nominal GDP to real GDP. Geometric mean annual real GDP per capita growth rates (using constant local currency units) come from WDI. Stock returns come from Datastream, where the Morgan Stanley Capital International (MSCI) total return indices with dividends being reinvested are used with CPI deflators from the World Bank's World Development Indicators (WDI). For annual real returns, inflation is measured from December to December. Returns for the BRIC countries (Brazil, Russia, India, and China) start after 1988 and their per capita real GDP growth rate is computed for the same years as for the stock returns.

Figure 2



Real per capita GDP growth rate per annum (on left in purple) and real equity return per annum (on right, in gray), 1988-2011. The real return data (dividends plus capital gains, adjusted for inflation, in local currency units) are from MSCI (2012). Real per capita GDP growth rates are from the World Bank. For the BRIC countries of Brazil, Russia, India, and China, the numbers start in 1993 or 1995 rather than 1988.

markets to assign higher P/E and price-to-dividend multiples when economic growth is expected to be high, which in turn means that companies must produce higher growth in earnings per share and dividends per share to meet investors' expectations and justify the current price. Unless companies achieve these increases by investing in positive-NPV projects, the higher prices paid will have the effect of reducing realized returns because more capital must be committed by investors to receive the same level of earnings and dividends.

Now, it's true that if earnings and dividend growth eventually turn out to be as high as expected, then overall shareholder returns will not be affected. But a variety of studies have reported that, when the dividend yields of U.S. companies are below their historical average, the growth rate of future dividends per share generally turns out to be lower, not higher, than the historical average growth rate.[7]

A second explanation for the negative correlation—one offered by Jeremy Siegel in his book *Stocks for the Long Run*—centers on the reality that, in many countries, the biggest publicly traded companies are multinationals that earn some or most of their earnings abroad. For example, the Finnish company Nokia sells only a small percentage of its products in Finland. International operations could plausibly lower the correlation between per capita GDP growth and stock returns. On the other hand, it is hard to see how international operations would cause a *negative* correlation.

The third explanation for the negative correlation between per capita GDP growth and stock returns—and in my view the most important—begins with the recognition that stock returns are determined not by growth in economy-wide earnings, but by improvement in measures of firm-specific corporate performance, such as growth in earnings *per share* and return on equity, that reflect *the amount of equity capital contributed by investors and the efficiency with which such capital is used.*

To make this point more clear, let's turn to the case of the U.S., where companies have returned large amounts of capital to investors through a combination of dividends and stock buybacks. During the period 1900-2011, the average earnings yield for U.S. companies has been just under 7%. At the same time, the average dividend yield has been about 4.2%. These two figures together imply that U.S. companies have reinvested cash flows that have annually averaged about 2.8% of their current market cap, which in turn suggests that the real growth rate of dividends per share should have been about 2.8%. At first glance, it seems puzzling that real dividends per share have not grown faster than the 1.3% per year reported in Table 1, unless companies on average have been consistently investing in negative NPV investments.

But this apparent puzzle disappears if we also take into account a bias in the inflation adjustments and two tendencies of U.S. companies: (1) large grants of employee stock options, especially among technology companies beginning in the 1980s; and (2) large payouts of corporate cash in cash-financed takeovers, beginning in the 1960s.

Quantitatively, the most important reason that the measured growth rate of real dividends per share is only 1.3% is because the inflation rate is overestimated by about 1% per year, which means that the true growth rate of real dividends per share has been about 2.3%.[8] Thus, the true discrepancy, once inflation is correctly measured, is between a 2.8% reinvestment rate and a 2.3% growth rate of real dividends per share.

The two tendencies of U.S. corporations that affect per share growth account for the rest of the discrepancy. The exercise of employee stock options has the effect of slowing the growth rate of real dividends per share by expanding the number of shares, and thereby diluting both EPS and dividends per share. Quantitatively, the bias is probably about 0.1% per year over the entire 1900-2011 period, since employee stock options were not common prior to the 1980s.[9]

In addition to paying out cash dividends, U.S. corporations have paid out large amounts of cash to shareholders by repurchasing shares.[10] The reduction in the number of shares outstanding affects the growth rate of dividends per share, so these share repurchases have already been accounted for in the growth rate of real dividends per share. But the dividend yield underestimates the average cash distribution to shareholders relative to the market value of equity for another reason. Cash used by one company to acquire another publicly traded company is equivalent to a share repurchase in distributing cash, but it doesn't affect the number of shares outstanding of the acquiring company. A cash-financed acquisition is merely using company A's cash to retire company B's shares.

In the last 50 years, a large amount of cash has been distributed in this manner. And over the entire 112-year sample period, cash-financed acquisitions per year probably average about 0.4% of the market value of equities. After adjusting for the effects of the understatement of inflation, employee stock option dilution, and cash-financed takeovers on cash payouts, the entire gap between the reported 1.3% growth rate of real dividends per share and the 2.8% reinvestment rate is accounted for.

Why are such distributions of capital important in explaining the high returns of U.S. companies? As noted earlier, the executives of public companies in all countries face political and social pressure to maintain or pursue corporate growth, even in cases where such growth is likely to produce less than competitive returns and reduce market values. Especially for large, established companies that are generating far more operating cash flow than they can profitably reinvest—think about GE or IBM—large annual corporate distributions in the form of dividends and stock repurchases play a critically important corporate governance function by helping managers to resist such pressures for growth.[11] Although often criticized

7. See John Campbell and Robert Shiller "Valuation Ratios and the Long-run Stock Market Outlook: An Update" (2001) and Robert Arnott and Clifford Asness "Surprise! Higher Dividends = Higher Earnings Growth," *Financial Analysts Journal* (2003).

8. *Toward A More Accurate Measure of the Cost of Living* (1996), also known as the Boskin Commission report, concludes that the CPI overstated U.S. inflation by about 1.3% per year from 1978–1996, with a smaller bias prior to 1978. Adjustments to the CPI computation made as a result of the report have reduced the bias since then, so that the average bias since 1900 is about 1.0% per year. An upward bias in the U.S. inflation rate implies that the average annual real return on U.S. stocks in Table 1 of 6.2% is also biased downwards. When real returns are measured in U.S. dollars, all countries would have an identical bias in their real returns, so the correlation of -0.32 reported in Table 1 would be unchanged.

9. Another way of thinking about such dilution is that, until the accounting rules were changed in 2002, U.S. companies that granted employee stock options were overstating their earnings because the employee stock options were not expensed.

10. Starting in 1984, U.S. companies began to pay out a substantial portion of earnings in the form of share repurchases. For evidence on the time series of aggregate cash payouts in the form of both dividends and share repurchases, see Harry DeAngelo, Linda DeAngelo, and Douglas J. Skinner, "Corporate Payout Policy," *Foundations and Trends in Finance* (2008).

in the popular business press as admissions of failure to find investment opportunities, such distributions of excess corporate cash and capital effectively force the managements of such companies to put the reinvestment decision back in the hands of their investors.[12] Without such payout policies, many companies have wasted massive amounts of investor capital in misguided attempts to maintain sales in declining businesses or get into unfamiliar ones. Think about corporate Japan, or the recent experience of companies like Eastman Kodak, the camera and film manufacturer whose business was decimated by digital photography and the replacement of cameras by smartphones.

In addition to the political and social pressures, there is a behavioral explanation of corporate managers' bias toward excessive retention and overinvestment, including overinvestment in acquisitions. Both successful—and unsuccessful—entrepreneurs and top corporate executives tend to be overly optimistic and confident about their own abilities. Managers who are prone to such excessive optimism are inclined to overinvest—that is, to take on projects that fail to earn their cost of capital—because of their habitual or instinctive tendency to emphasize what can go right while downplaying potential downsides.[13]

In sum, although higher capital investment by companies generally means higher growth rates for national economies (at least over the near term), it is by no means a reliable prescription for higher returns to shareholders over the longer term. In the U.S., for example, some industries have consistently invested in negative NPV projects, causing significant losses for their shareholders. Industries that have experienced remarkable growth during the last century include airlines, automobiles, computer hardware and software, and pharmaceuticals. At the same time, industries such as railroads, steel, and tobacco have declined sharply in relative importance. But the shareholders of airlines have not gotten rich, nor have the owners of auto companies during the last 45 years. Instead, investors in these industries have seen many billions of dollars wasted in value-destroying negative-NPV projects. Tobacco companies, on the other hand, have done very well for their shareholders, despite hundreds of billions of dollars paid out to settle lawsuits, in part by paying out large fractions of their still considerable operating cash flows in the form of dividends.

But perhaps the most compelling evidence of the importance of corporate payout policy comes from what was once the largest of U.S. industries in terms of market value. At the beginning of 1900, railroads made up 63% of the market cap of U.S. stocks—a number that, by 2002, had fallen to less than 0.2% of the total U.S. market cap.[14] Much as happened in the U.S. auto, steel, and airline industries, after initial periods of growth and profitability, the returns on massive amounts of capital that were reinvested (instead of being paid out) by U.S. railroads later proved to be disappointingly low or even negative, destroying large amounts of shareholder value.

The Sources of Economic Growth

There is a huge literature on the determinants of economic growth. Although this article will not attempt to do more than sketch the broad outlines of this work, in so doing I will try to emphasize the connection, or lack thereof, between the determinants of growth and stock returns.

Simply put, economic growth results mainly from increases in three main inputs: labor, capital, and technology. The efficiency with which these inputs are used also matters, and such efficiency is affected by a nation's culture, institutions, and government policy.

Increases in labor inputs come from increases in the general population, in the fraction of the population that is able and willing to work, and in the human capital of the workforce. In almost all developed and developing countries, the non-agricultural labor force has become a larger fraction of the population over time as the adult children of what was once the largest class of workers, subsistence farmers, have moved to urban areas and taken manufacturing and service jobs. In much of Europe and its offshoots of Australia, Canada, New Zealand, and the U.S., this transition took place very gradually. In East Asia, by contrast, this transition has been occurring with remarkable speed, providing a major impetus to growth.

Another source of increased labor is, somewhat paradoxically, the drop in birth rates that has been taking place in most of the world. The paradox of a decline in birth rates leading to higher labor force growth is attributable to two effects. Most obviously, lower birth rates provide an opportunity for women who might otherwise be caring for children to enter the paid work force. Less obviously, the large numbers of children who were born when birth rates were still high enter the labor force 20 years later, but do not retire for another 40 years. Starting 20 years after birth rates have started to fall, there are both relatively few retirees and relatively few

11. Another reason that GDP growth does not necessarily translate into high returns for minority stockholders, with particular relevance for countries outside the U.S., is that managers may expropriate profits through sweetheart deals, tunneling, and other ruses. There is a large literature focusing on this, but most of its emphasis has been on how corporate governance problems would keep public equity markets from becoming large. The assumption is that minority investors would correctly evaluate in advance the chance of receiving future dividends, and if the legal and institutional mechanisms are weak, firms would be unable to sell equity to the public at terms that are attractive enough to make it an optimal financing/ownership mechanism. This assumes that investors price protect themselves. If investors do not price protect themselves, then it is possible that public equity markets would be bigger than otherwise, but that realized returns would be low because profits would accrue to managers rather than minority shareholders.

12. One particularly instructive example of the importance of such distributions are energy master limited partnerships, which routinely pay out as much as 90% of their operating cash flow, only to get most of that capital back through follow-on equity offerings. During the 30 months from mid-2008 through 2010, energy MLPs paid out an estimated $18 billion to their unitholders while raising $16 billion in follow-on offerings—and from essentially *the same* group of investors.

13. See J.B. Heaton, "Managerial Optimism and Corporate Finance," *Financial Management* (Summer 2002).

14. According to Dimson, Marsh, and Staunton, *Triumph of the Optimists* (2002).

children, thus making the fraction of the population in their prime working years unusually high for a period of about 40 years. In cases where the drop in birth rates takes place suddenly, as has happened in many East Asian countries, this "demographic dividend" has supercharged growth rates.

Yet another source of increased labor inputs is increased human capital per worker. An increase in human capital can result from improvements in health, but the more important source of increase has been through increased levels of education. Not all education is the same, of course. It is widely believed, for example, that engineering and technical education has a positive effect on economic growth, while the effect of adding to the supply of lawyers in the United States is less clear.

Along with increased inputs of labor, infusions of new capital and the associated increases in capital per worker can also lead to higher economic growth. Although capital can be accumulated in a number of different ways, the most fundamental source is the savings of individuals. Apart from amounts invested in housing or small private enterprises, such savings tend to be channeled into an economy through two main conduits: (1) governments, particularly through financial institutions that are owned or controlled by the public sector; and (2) private-sector banks and corporations, which increase their own capital and investment through the issuance of new securities and/or the retention and reinvestment of earnings.

In analyzing differences in economic growth rates, it's useful to start by looking at the well-known critiques of Asia's economic miracle by Paul Krugman and Alwyn Young.[15] In their widely cited articles, Krugman and Young argue that the high growth rates achieved by the Soviet Union during the period 1930-1970, and by many East Asian countries from 1960-1993, resulted mainly from taking economies with vast supplies of under-utilized labor but very little capital, and then bringing together capital (from high savings rates) and labor (by moving people out of subsistence agriculture) in combination with mainly imported technology. While this transition was occurring, these economies experienced exceptionally high rates of economic growth, bolstered by the demographic dividend that is partly responsible for China's current high rate of growth.

But while I agree with Krugman and Young that much of the economic growth in emerging markets is attributable to high savings rates (with a modest role for foreign direct investment) and the more efficient use of labor, I want to emphasize that even a continuing increase in the supply of these two inputs—which, as Krugman and Young suggest, is itself a doubtful proposition—is not likely to translate into higher per share profits for the shareholders of their listed companies. As we have already seen, stock returns tend to be high when corporate earnings are reinvested in positive-NPV projects, which results in a high growth rate of dividends per share. If uninterrupted growth is the paramount objective of a national or corporate policy, companies can still grow their top and (even their) bottom lines by reinvesting in negative-NPV projects—and countries can ramp up their growth rates—while inflicting losses on shareholders. But over the longer term, the failure to provide investors with adequate returns on capital is likely to reduce economic growth. Again, think about the case of Japan during the last 20 years.

In addition to increased inputs of capital and labor, economic growth also comes from technological progress, as inputs are transformed into outputs more efficiently. But much of the efficiency benefits of technological change end up accruing not to investors, but to consumers in the form of lower prices and higher-quality products, as competition between companies limits the ability to boost profit margins when costs decline.[16] Let me illustrate this point with two examples: the agricultural industry and the airline industry.

One hundred and fifty years ago, roughly 90% of the labor force in Europe and North America worked in agriculture. Thanks mainly to technological advances (such as improved seeds) and increased capital (synthetic fertilizer, tractors, etc.), agricultural output per farmer has skyrocketed, and today only a few percent of the population in developed countries work in the agriculture sector. But have the owners of farmland gotten rich? The answer is no, or at least not as a result of increases in farming profits rather than government subsidies. The increase in agricultural output has been so vast that the benefits have accrued almost entirely to the consumers of food. Standards of living have improved because of the vast number of workers who now produce output in other sectors of the economy instead of the agricultural sector.

Or think about the effects of technological change on a company in the airline industry, Delta Airlines. Over the last 60 years, improvements in airplanes, such as the replacement of propeller aircraft by jets and more efficient jet engines, have permitted Delta to make dramatic reductions in the inflation-adjusted costs of flying its passengers over long distances. Furthermore, modern computerized airline reservations systems have allowed Delta to charge different passengers different prices for seats on the same flight, a practice known as "yield management," and thereby maximize the average ticket price while also filling a high percentage of seats. If Delta was the only airline with lower costs and higher revenue per passenger, it would be able to boost its profit margins. But since other competing airlines have also benefited from

15. See Paul Krugman, "The Myth of Asia's Economic Miracle," *Foreign Affairs* (1994) and Alwyn Young, "The Tyranny of Numbers: Confronting the Statistical Realities of the East Asian Growth Experience," *Quarterly Journal of Economics* (1995). Krugman's article gives a non-technical summary of Young's research. Because of the difference in the speed of publication between *Foreign Affairs* and the *Quarterly Journal of Economics*, Krugman's article was published first, even though Young's article was written first.

16. For discussions of why technological change does not necessarily benefit shareholders, see Warren Buffett (1999) in *Fortune* and Jeremy Siegel (2000) in the *Wall Street Journal*.

improved technology, competition has driven down the average ticket price, and Delta's shareholders have not earned high returns. Indeed, Delta has joined Air Canada, United, US Airways, American, Continental, TWA, Pan Am, defunct Belgian carrier Sabena, and countless other airlines around the world in declaring bankruptcy one or more times.

Predicting Future Returns

In general, there is no consensus about how to estimate future stock returns. This is especially true for emerging markets, where frequently there are only limited data on past stock returns. Limited historical information on stock returns is not a constraint, however, since such data are irrelevant for estimating future returns, whether in emerging markets or developed countries. For estimating future returns, forward-looking information is needed. This point has been made before, although possibly not as explicitly, by Jeremy Siegel (in his 2002 book) and by Gene Fama and Ken French, among others.[17] Here I go one step farther and argue that knowledge of the future real growth rate for an economy, even if knowable in advance, is also largely irrelevant. My argument thus suggests that whether the Chinese economy ends up growing by 7% per year, or by 3% per year, for the foreseeable future is unimportant for the future returns on Chinese stocks.

In what follows, I argue that one needs only four pieces of information to estimate future equity returns. The first is the current P/E ratio, although earnings must be smoothed to adjust for business cycle fluctuations. The second is the fraction of corporate profits that will be paid out to shareholders in the form of share repurchases and dividends. The third is the return on capital for the reinvested earnings. As already noted, if the money is invested in positive-NPV projects, a high P/E ratio can be justified. The fourth is the probability of catastrophic loss—that is, the chance that "normal" profits are an upwardly biased measure of expected profits because of "tail risks" stemming from the possibility of low-probability, large-loss events.

To see why future economic growth is largely irrelevant to predicting stock returns in an economy, it helps to start by recognizing that investors realize returns only on the shares that they hold, not on shares that may later be issued by the same companies to other investors. And this in turn implies that the returns on existing shares will be abnormally high only if a corporation's earnings are reinvested in projects with higher returns than the market had expected. Part of an economy's growth, as we have already seen, can be attributed to savings invested in new companies, and to the issuance of new securities by existing companies. But the gains on this capital investment do not necessarily accrue to today's shareholders.

In the short run, of course, there is ample evidence that unexpected changes in economic growth affect stock prices. Stock prices fall when the probability of an economic recession increases, and prices rise when the probability of economic recovery increases. Recessions are definitely bad for corporate profitability, and cyclical recoveries are good. But while such cyclical effects clearly have an effect on equity valuations, the effects should be largely transitory, mainly because they typically do not have a big impact on the present value of the earnings and dividends of a given company.[18]

What about the possibility that today's stock prices are depressed by general concern that a catastrophic event may wipe out a country's financial markets? This would show up in both a high promised yield on bonds, and depressed P/E ratios. In this scenario, the earnings yield on stocks will overestimate future expected equity returns for the same reason that the yield to maturity on corporate bonds overestimates the expected return. In both cases, there is a "default" probability, and the expected returns are lower than the "promised" returns.

This is a reasonable characterization, at least in hindsight, of how stock and bond returns looked to many (if not most) investors during the panic of 2008. But now let's move to today's market conditions, with the S&P 500 around 1400 and the Dow over 13,000. If past stock returns are irrelevant for predicting future stock returns, and future economic growth rates are also irrelevant, what is likely to matter?

The answer is fairly straightforward: earnings yields. Following Jeremy Siegel and using a formula that has become known as the "Shiller earnings yield," one can forecast future compounded real stock returns as follows: $E(r) = E^*/P$, where E^* is *normalized* earnings per share (that is, EPS smoothed to take out business cycle effects).[19]

As we have already discussed at some length, corporate earnings can either be paid out or reinvested (in capital investments or acquisitions). But as long as we assume that companies earn their required rate of return on reinvested capital, the compounded real return will not be affected by whether earnings are paid out or reinvested.

Of course, P/E ratios fluctuate all the time, and such fluctuations can be attributed to changes in either the numerator or the denominator. Since current earnings fluctuate based on business cycle effects, a market P/E could be temporarily high because earnings are temporarily depressed. This is why Siegel recommends the use of "smoothed" estimates of earnings that try to remove the effects of the business cycle.

In a 2001 study, John Campbell and Robert Shiller use

17. See Eugene Fama and Kenneth French "The Equity Premium," *Journal of Finance* (2002).

18. I believe that the large stock price effects associated with recessions are partly due to increases in risk aversion at the bottom of a recession, but also partly due to an irrational overreaction. During the 2008 financial panic, for example, drops in stock prices can be attributed to three factors: (1) lower expected cash flows, due to an increase in the possibility of a worldwide depression; (2) higher risk, due to a higher probability of extreme scenarios, and (3) greater risk aversion, which corresponds to a higher market price per unit of risk. The third point results in higher expected returns on a point-forward basis. Irrational overreaction would occur if cash flow forecasts became excessively pessimistic or perceptions of risk were higher than objectively justified. Overreaction results in excessive volatility and mean reversion over multi-year horizons.

19. See Jeremy Siegel, *Stocks for the Long Run* (2008, Chapter 7).

a 10-year moving average of earnings on the S&P 500 to smooth out the effects of the business cycle.[20] Dividing this moving average of earnings by the current level of the S&P 500 index provides what has come to be known as the Shiller earnings yield on the market. Campbell and Shiller report finding that when smoothed earnings yields are lower than historical averages (i.e., when P/E ratios are high), future returns also tend be lower than average. In other words, P/E ratios tend to revert toward a mean, but more often than not through changes in price rather than changes in earnings.

Conclusion

Over long periods of time, the cross-country correlation of per capita real GDP growth and real stock returns has been negative. This pattern has been true for both developed countries and emerging markets, and whether returns are measured in local currencies or U.S. dollars. While historical performance, as the saying goes, is no guarantee of future returns, the evidence flies in the face of the intuition that economic growth should benefit stockholders.

The most plausible explanation of this finding is that consumers and workers are the primary beneficiaries of economic growth, and not the shareholders of existing companies. This finding, however, does not mean that companies should not aim for continuous improvement in their technology and, indeed, all aspects of their business. If the competition is becoming more efficient, failing to keep up with competitors will result in lower profits.

But for corporate managements, the key to adding value is investing operating cash flow in all available positive-NPV projects, while *at the same time* returning any excess cash and capital to investors through dividends and stock repurchases. A rapidly growing economy may result in a tendency for companies to overinvest, perhaps out of a fear of losing market share. As ample evidence from the corporate finance literature suggests, this kind of overinvestment—and the temporary economic growth it produces—does not benefit the shareholders of the existing companies.

JAY RITTER is Cordell Professor of Finance at the University of Florida. He is best known for his research on initial public offerings, but he has published articles on many other topics as well.

References

Arnott, Robert D., and Clifford S. Asness, 2003, Surprise! Higher Dividends = Higher Earnings Growth, *Financial Analysts Journal* 59, No. 1, 70-87.

Boskin, Michael, Ellen R. Dulberger, Robert J. Gordon, Zvi Griliches, and Dale Jorgenson, 1996, *Toward A More Accurate Measure of the Cost of Living.*

Buffett, Warren, 1999, Mr. Buffet on the Stock Market, *Fortune* November 22, 1999.

Campbell, John Y., and Robert J. Shiller, 2001, Valuation Ratios and the Long-run Stock Market Outlook: An Update, Cowles Foundation Discussion Paper No. 1295.

DeAngelo, Harry, Linda DeAngelo, and Douglas J. Skinner, 2008, Corporate Payout Policy, *Foundations and Trends in Finance*, 95-297.

Dimson, Elroy, Paul Marsh, and Mike Staunton, 2002, *Triumph of the Optimists: 101 Years of Global Investment Returns*, Princeton: Princeton University Press.

Dimson, Elroy, Paul Marsh, and Mike Staunton, 2005, *ABN AMRO Global Investment Returns Yearbook 2005.*

Dimson, Elroy, Paul Marsh, and Mike Staunton, 2010, *Credit Suisse Global Investment Returns Yearbook 2010.*

Dimson, Elroy, Paul Marsh, and Mike Staunton, 2012, *Credit Suisse Global Investment Returns Sourcebook 2012.*

Elroy Dimson, Paul Marsh, and Mike Staunton, "Global Evidence on the Equity Risk Premium," *Journal of Applied Corporate Finance* 15 (Fall 2003), 27-38.

Fama, Eugene F., and Kenneth R. French, 2002, The Equity Premium, *Journal of Finance* 57, 637-659.

Heaton, J.B., 2002, Managerial Optimism and Corporate Finance, *Financial Management* (Summer), 33-45.

Krugman, Paul, 1994, The Myth of Asia's Economic Miracle, *Foreign Affairs* 73, 62-78.

Maddison, Angus, 1995, *Monitoring the World Economy 1820-1992* Paris: OECD Development Centre Studies.

Maddison, Angus, 2010 his website.

Ritter, Jay R., 2005, Economic Growth and Equity Returns, *Pacific-Basin Finance Journal* 13, 489-503.

Siegel, Jeremy J., 2000, Big Cap Tech Stocks Are a Sucker Bet, *Wall Street Journal,* March 14, A14.

Siegel, Jeremy J, 2002, *Stocks for the Long Run* (Third edition) McGraw-Hill.

Siegel, Jeremy J., 2008, *Stocks for the Long Run* (Fourth edition) McGraw-Hill.

World Bank, 2012, *World Development Indicators* Washington, D.C.

Young, Alwyn, 1995, The Tyranny of Numbers: Confronting the Statistical Realities of the East Asian Growth Experience, *Quarterly Journal of Economics* 110, 641-680.

20. Robert Shiller's website maintains an updated Excel file with the smoothed earnings yield on the S&P 500 index.

Appendix **Table A-1 Levels and Growth Rate of Per Capita GDP for 19 Countries, 1900-2011**

Country	Real per capita GDP, $1990 1900	2011	Per annum	Population in 1900, m	Population growth Cumulative	Per annum
United Kingdom	4,492	22,866	1.48%	38.000	65%	0.45%
New Zealand	4,298	18,000	1.30%	0.967	357%	1.38%
United States	4,091	30,755	1.85%	76.212	309%	1.28%
Australia	4,013	25,406	1.68%	4.000	467%	1.57%
Switzerland	3,833	24,985	1.70%	3.525	124%	0.73%
Belgium	3,731	23,309	1.66%	6.136	79%	0.52%
Netherlands	3,424	24,131	1.78%	5.616	197%	0.99%
Denmark	3,017	23,377	1.86%	2.182	157%	0.86%
Germany	2,985	21,175	1.78%	56.000	46%	0.34%
Canada	2,911	25,104	1.96%	5.500	527%	1.66%
France	2,876	21,891	1.85%	41.000	54%	0.39%
Ireland	2,736	25,304	2.30%	4.466	3%	0.03%
Sweden	2,209	24,941	2.21%	5.140	83%	0.54%
Norway	1,877	27,560	2.45%	2.240	123%	0.72%
Spain	1,786	18,808	2.14%	20.750	123%	0.72%
Italy	1,785	18,940	2.15%	33.000	84%	0.55%
Finland	1,668	23,449	2.41%	2.656	103%	0.64%
South Africa	1,602	4,830	1.13%	5.014	907%	2.10%
Japan	1,180	22,333	2.69%	42.000	205%	1.01%

Sources: For the real per capita GDP numbers, "Statistics on World Population, GDP and Per Capital GDP, AD 1-2008" (horizontal file, copyright Angus Maddison, University of Groningen) available at http://www.ggdc.net/maddison/oriindex.htm, in 1990 international Geary-Khamis (purchasing power parity-adjusted) dollars. Ireland and South Africa have 1913 numbers rather than 1900 numbers for real per capita GDP, so the per annum growth rate of real GDP per capita is computed by taking the 98th root of the 2011/1913 ratio. The 2011 numbers come from taking the 2008 Maddison numbers and multiplying by the ratio of 2011 to 2008 real GDP per capita in local currency unit numbers from the World Bank. For Finland and New Zealand, tradingeconomics.com is the source of the 2011 real per capita GDP numbers relative to 2008.

Population in 1900 is given in millions, with 1900 populations from http://en.wikipedia.org/wiki/List_of_countries_by_population_in_1900 except for South Africa, Finland, France, and Ireland. The Irish population is from www.libraryireland.com, which gives a U.K population of 41.150 million in 1900. The Finnish population is from http://www.vaestoliitto.fi/@Bin/236655/YB+09_Statistics.pdf for 1900. The French population in 1900 is given as 38 million by Wikipedia but 41 million at http://www.worldmapper.org/posters/worldmapper_map9_ver5.pdf.

http://en.wikipedia.org/wiki/South_Africa gives a South African population of 5.014 million. 2011 populations are from the Population Reference Bureau at http://www.prb.org/pdf11/2011population-data-sheet_eng.pdf .

Journal of Applied Corporate Finance (ISSN 1078-1196 [print], ISSN 1745-6622 [online]) is published quarterly, on behalf of Morgan Stanley by Wiley Subscription Services, Inc., a Wiley Company, 111 River St., Hoboken, NJ 07030-5774. Postmaster: Send all address changes to JOURNAL OF APPLIED CORPORATE FINANCE Journal Customer Services, John Wiley & Sons Inc., 350 Main St., Malden, MA 02148-5020.

Information for Subscribers *Journal of Applied Corporate Finance* is published in four issues per year. Institutional subscription prices for 2012 are:

Print & Online: US$484 (US), US$580 (Rest of World), €375 (Europe), £299 (UK). Commercial subscription prices for 2012 are: Print & Online: US$647 (US), US$772 (Rest of World), €500 (Europe), £394 (UK). Individual subscription prices for 2011 are: Print & Online: US$109 (US), £61 (Rest of World), €91 (Europe), £61 (UK). Student subscription prices for 2012 are: Print & Online: US$37 (US), £21 (Rest of World), €32 (Europe), £21 (UK).

Prices are exclusive of tax. Australian GST, Canadian GST and European VAT will be applied at the appropriate rates. For more information on current tax rates, please go to www.wileyonlinelibrary.com/tax-vat. The institutional price includes online access to the current and all online back files to January 1st 2008, where available. For other pricing options, including access information and terms and conditions, please visit www.wileyonlinelibrary.com/access

Journal Customer Services: For ordering information, claims and any inquiry concerning your journal subscription please go to www.wileycustomerhelp.com/ask or contact your nearest office.
Americas: Email: cs-journals@wiley.com; Tel: +1 781 388 8598 or +1 800 835 6770 (toll free in the USA & Canada).
Europe, Middle East and Africa: Email: cs-journals@wiley.com; Tel: +44 (0) 1865 778315.
Asia Pacific: Email: cs-journals@wiley.com; Tel: +65 6511 8000.
Japan: For Japanese speaking support, Email: cs-japan@wiley.com; Tel: +65 6511 8010 or Tel (toll-free): 005 316 50 480.
Visit our Online Customer Get-Help available in 6 languages at www.wileycustomerhelp.com

Production Editor: Joshua Gannon (email:jacf@wiley.com).
Delivery Terms and Legal Title Where the subscription price includes print issues and delivery is to the recipient's address, delivery terms are Delivered Duty Unpaid (DDU); the recipient is responsible for paying any import duty or taxes. Title to all issues transfers FOB our shipping point, freight prepaid. We will endeavour to fulfil claims for missing or damaged copies within six months of publication, within our reasonable discretion and subject to availability.

Back Issues Single issues from current and recent volumes are available at the current single issue price from cs-journals@wiley.com. Earlier issues may be obtained from Periodicals Service Company, 11 Main Street, Germantown, NY 12526, USA. Tel: +1 518 537 4700, Fax: +1 518 537 5899, Email: psc@periodicals.com

This journal is available online at Wiley Online Library. Visit www.wileyonlinelibrary.com to search the articles and register for table of contents e-mail alerts.

Access to this journal is available free online within institutions in the developing world through the AGORA initiative with the FAO, the HINARI initiative with the WHO and the OARE initiative with UNEP. For information, visit www.aginternetwork.org, www.healthinternetwork.org, www.healthinternetwork.org, www.oarescience.org, www.oarescience.org

Wiley's Corporate Citizenship initiative seeks to address the environmental, social, economic, and ethical challenges faced in our business and which are important to our diverse stakeholder groups. We have made a long-term commitment to standardize and improve our efforts around the world to reduce our carbon footprint. Follow our progress at www.wiley.com/go/citizenship

Abstracting and Indexing Services
The Journal is indexed by Accounting and Tax Index, Emerald Management Reviews (Online Edition), Environmental Science and Pollution Management, Risk Abstracts (Online Edition), and Banking Information Index.

This journal is printed on acid-free paper.

Energy Policy 133 (2019) 110891

Contents lists available at ScienceDirect

Energy Policy

journal homepage: www.elsevier.com/locate/enpol

Regulated equity returns: A puzzle

David C. Rode[a,*], Paul S. Fischbeck[a,b]

[a] Department of Social and Decision Sciences, Carnegie Mellon University, Pittsburgh, PA, 15213, USA
[b] Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA, 15213, USA

ARTICLE INFO

Keywords:
Discount rate
Electric utility
Rate of return regulation
Valuation

ABSTRACT

Based on a database of U.S. electric utility rate cases spanning nearly four decades, the returns on equity authorized by regulators have exhibited a large and growing premium over the riskless rate of return. This growing premium does not appear to be explained by traditional asset-pricing models, often in direct contrast to regulators' stated intent. We suggest possible alternative explanations drawn from finance, public policy, public choice, and the behavioral economics literature. However, absent some normative justification for this premium, it would appear that regulators are authorizing excessive returns on equity to utility investors and that these excess returns translate into tangible profits for utility firms.

1. Introduction

In economics, the equity-premium puzzle refers to the empirical phenomenon that returns on a diversified equity portfolio have exceeded the riskless rate of return on average by more than can be explained by traditional models of compensation for bearing risk. Since Mehra and Prescott's (1985) initial paper on the subject, a large body of research has attempted to explain away the puzzle, but without much success (Mehra and Prescott, 2003). The most likely explanations for the premium appear to reside outside of classical equilibrium models. We call the reader's attention to the Mehra-Prescott puzzle as a means of introducing our instant problem, of which it may be considered an applied case. Simply put: why are the equity returns authorized by electric utility regulators so high, given that riskless rates are so low?

Our scope is as follows. We employ a much larger dataset than has previously been examined in the literature and seek to explain the rates of return *authorized* by state electric utility regulators. We investigate the extent to which the actual returns authorized can be explained by the Capital Asset Pricing Model (CAPM), which regulators (and others) purport to use. We also examine whether the CAPM is capable of explaining the clear trend of rising risk premiums present over the last four decades in electric utility rate cases.

While previous studies have investigated rates of return for regulated electric utilities, the majority of this work has either examined *actual* rates of return to utility stockholders, relied on very limited samples of rate cases, or tested a variety of hypotheses connecting utility earnings to various structural and institutional factors. Table 1 summarizes the previous literature most similar to our study. By contrast, our study employs a far larger sample of rate cases (1,596) than previously examined in the literature. In addition, our focus on authorized rates of return highlights the impact of regulatory rate-setting on consumers, as opposed to stockholders, to the extent that authorized rates are used to set utility revenue requirements, while earned returns accrue to stockholders. This setting also enables us to analyze rate-setting in the context of regulatory decision-making. Actual rates of return earned by utilities can differ from the rates of return authorized by regulators due to factors such as the impact of weather on demand, but primarily due to the operational performance of a utility, including its ability to operate efficiently and control costs to those approved by regulators.

This regulated equity return puzzle is important not just from a theoretical asset-pricing perspective, but also for very practical reasons. The database used in this study reflects more than $3.3 trillion (in 2018 dollars) in cumulative rate-base exposure.[1] An error or bias of merely one percentage point in the allowed return would imply tens of billions of dollars in additional cost for ratepayers in the form of higher retail power prices and could play a profound role in the allocation of investment capital. Coupled with utilities' tendencies toward excessive capital accumulation under rate regulation (Averch and Johnson, 1962; Spann, 1974; Courville, 1974; Hayashi and Trapani, 1976; Vitaliano

* Corresponding author.
E-mail addresses: rode@andrew.cmu.edu (D.C. Rode), fischbeck@cmu.edu (P.S. Fischbeck).

[1] This figure reflects the simple cumulative sum of authorized rate bases across all cases. Because rate-base decisions may remain in place for several years, this sum most likely underestimates the actual figure, which should be the authorized rate base in each year examined, whether or not a new case was decided. We cite this figure merely as evidence of the substantial magnitude of the costs at stake.

https://doi.org/10.1016/j.enpol.2019.110891
Received 22 January 2019; Received in revised form 18 July 2019; Accepted 19 July 2019
Available online 25 July 2019
0301-4215/

Table 1
Previous studies of the determinants of electric utility rates of return.

Study	Sample	Description
Joskow (1972)	20 cases in New York between 1960 and 1970	Only capital markets parameter included was cost of debt. Focused on the requested rate of return.
Joskow (1974)	174 cases between 1958 and 1972	No CAPM parameters tested. Regulators tended to ignoring overearning as long as prices were falling.
Hagerman and Ratchford (1978)	79 survey responses from utilities about their last rate case	Used authorized rates. Found positive coefficients related to beta and the debt/equity ratio.
Roberts et al. (1978)	59 cases from 4 Florida utilities between 1960 and 1976	No CAPM parameters tested. Only structural factors examined.
Roll and Ross (1983)	Utility stock returns between 1925 and 1980	No authorized returns used. CAPM underestimates returns relative to the APT.
Pettway and Jordan (1987)	58 electric service companies between 1969 and 1976	Used stockholder returns only.
Binder and Norton (1999)	92 firms	Used stockholder returns to estimate beta. Suggested that regulation causes cash flow “buffering” and that firms may be underearning.
PJM Interconnection (2016)	22 regulated firms between 2000 and 2015	Examined stockholder returns and found regulated firms had positive alpha.
Haug and Wieshammer (2019)	N/A	Regulators in continental Europe “uniformly adopt the [CAPM]” and courts have ruled that the authorized rates are too low. The opposite finding to our study.

and Stella, 2009), the magnitude of the problem makes it incumbent on the industry and regulators to get it right.

There are also policy implications for market design and regulation. A recent PJM Interconnection (2016) study compared and contrasted entry and exit decisions in competitive and regulated markets to evaluate the efficiency of competitive markets for power. One finding that emerged from the study was that regulated utilities appeared to be “overearning” and had generated positive alpha, while competitive firms had not generated positive alpha.[2] Although the study used a limited time window of rate case data and focused on utility stock returns, not returns authorized by regulators, its findings are consistent with those we explore in more detail here.

As an old joke goes, an economist is someone who sees something work in practice and asks whether it can work in theory. Undoubtedly, the utility sector has been successful in attracting capital over the past four decades. We cannot necessarily say, however, that had returns been consistent with the dominant theoretical model used (and thus lower), this would still have been the case. Accordingly, this article also raises the question of whether our theoretical models of required return and asset pricing must be refined. Or, at the very least, whether there are important considerations that must be accounted for in the application of those models to the regulated electric utility industry.

In this article, therefore, we examine the historical data on authorized rates of return on equity in U.S. electric utility rate cases. We compare these rates of return to several conventional benchmarks and the classical theoretical asset-pricing model. We demonstrate that the spread between authorized equity returns (and also earned equity returns) and the riskless rate has grown steadily over time. We investigate whether this growing spread can be explained by classical asset-pricing parameters and conclude that it cannot. We then evaluate possible explanations outside of classical finance to suggest fruitful paths for future research. Specifically, we investigate whether the addition of variables for commission selection and case adjudication contribute explanatory power, in line with existing theories in the pubic choice literature. We conclude with a discussion of the policy implications of the observed premiums and how regulatory rate-setting could be adjusted to mitigate higher premiums.

Section 2 reviews the legal, regulatory, and financial foundations of rate of return determination for utilities. Section 3 describes the data used in our analysis and defines the risk premium on which our analysis is based. Section 4 presents the results of our analysis and outlines the various factors explored, including both classical financial factors and factors outside of the classical paradigm. Section 5 highlights the policy implications of our research, suggests potential mitigating strategies, and concludes.

2. Regulated equity returns and the Capital Asset Pricing Model

At the outset, let us make clear that we are addressing only *regulated* rates of return on equity in this article. We draw no conclusions or inferences about the behavior of returns on non-regulated assets. Our focus is limited to regulated returns because in such cases it is regulators who are tasked with standing in for the discipline of a competitive market and ensuring that returns are just and reasonable. For more than a century, U.S. courts have ruled consistently in support of this objective, while recognizing that achieving it requires consideration of numerous factors that are subject to change over time. The task set to regulators, then, is to approximate what a competitive market would provide, *if one existed.*

Mindful of this mandate, two U.S. Supreme Court decisions are commonly thought to provide the conceptual foundation for utility rate-of-return regulation. In *Bluefield Water Works & Improvement Co. v. Public Service Commission of West Virginia* (262 U.S. 679 (1923)), the Court identified eight factors that were to be considered in determining a fair rate of return, ruling that “[t]he return should be reasonable, sufficient to assure confidence in the financial soundness of the utility, and should be adequate, under efficient and economic management, to maintain and support its credit and enable it to raise money necessary for the proper discharge of its public duties.” This position was made more concrete in *Federal Power Commission v. Hope Natural Gas Company* (320 U.S. 591 (1944)), wherein the Court ruled that the “return to the equity owner should be commensurate with returns on investments in other enterprises having corresponding risks.”

In both *Bluefield* and *Hope*, the Court sought to balance the need for utilities to attract capital sufficient to discharge their duties with the need for regulators to protect ratepayers from what would otherwise be rent-seeking monopolists. These efforts in determining “just and reasonable” returns received significant assistance in the 1960s when groundbreaking advances in asset-pricing theory were made in finance. Specifically, the development of the Capital Asset Pricing Model (CAPM) (Sharpe, 1964; Lintner, 1965; Mossin, 1966) provided a rigorous framework within which the question of the appropriate rate of return could be addressed in an objective fashion. The security market line representation of the CAPM [1] set out the equilibrium rate of return on equity, r_E, as the sum of the rate of return on a riskless asset,

[2] In asset pricing models, positive alpha is evidence of non-equilibrium returns, meaning that investors are receiving compensation in excess of what would be required for bearing the risks they have assumed.

r_f, and a premium related to the level of risk being assumed that was defined in relation (through the factor β) to the expected excess rate of return on the overall market for capital, r_m.

$$r_E = r_f + \beta(r_m - r_f) \tag{1}$$

It is outside of the scope of this paper to delve too deeply into the foundations of asset pricing. We note, also, that the CAPM methodology is not the sole candidate for rate-of-return determination in utility rate cases. Morin (2006, p. 13) identifies four main approaches used in the determination of the "fair return to the equity holder of a public utility's common stock," of which the CAPM is but one.[3] Nevertheless, the concept of the appropriate rate of return on equity being a combination of a riskless rate of return and a premium for risk-bearing has since become widely accepted as a means of determining the appropriate authorized return on equity in state-level utility rate cases (Phillips, 1993, pp. 394–400). In contrast, the Federal Energy Regulatory Commission relies exclusively on the DCF approach, which is also common with natural gas utilities. For electric utilities, however, the CAPM in particular is seen as the "preferred" (Myers, 1972; Roll and Ross, 1983, p.22) and "most widely used" (Villadsen et al., 2017, p. 51) method in regulatory proceedings. Multi-factor approaches such as Arbitrage Pricing Theory (APT) (Ross, 1976) and the Fama and French (1993) framework are used with significantly less frequency in practice (Villadsen et al., 2017, p. 206). In other words, our focus on the CAPM is not solely because of its perceived normative status, but also because it is the method most regulators *say they are using.*

In *Hope,* however, the Court also advocated the "end results doctrine," acknowledging that regulatory methods were (legally) immaterial so long as the end result was reasonable to the consumer and investor. In other words, there was no single formula for determining rates. A typical example of the latitude granted by the doctrine is found in Pennsylvania Public Utility Commission (2016, p. 17): "The Commission determines the [return on equity] based on the range of reasonableness from the DCF barometer group data, CAPM data, recent [returns on equity] adjudicated by the Commission, and **informed judgment** [emphasis added]." Rate determination in practice is often not simply a matter of arithmetic; rather, it is an act of judgment performed by regulators. As a result, our investigation examines not just the relation of authorized rates to those implied by the CAPM, but also the potential for that relationship to be influenced by regulator judgment.

Before we turn to the data, however, let us dispense with an alternate formulation of the underlying question. In questioning the size of the premium and why equity returns are so high, one might also ask instead why the riskless rate is so low. Indeed, Mehra and Prescott (1985) ask this very question, before dismissing it on theoretical grounds. We revisit this question in light of recent data and ask whether the premium during the period in question is more a function of riskless rates being forced down by the Federal Reserve's intervention, than of equity premiums increasing (since the manifest intent of quantitative easing was to lower riskless rates).[4] Our historical data, as Section 3 indicates, do not support that hypothesis. The premium growth has persisted since the beginning of our data series in 1980 and has persisted across a variety of monetary and fiscal policy regimes.

3. Regulated electric utility returns on equity, 1980–2018

3.1. Historical authorized return on equity data

The data used in this study were collected and maintained by Regulatory Research Associates (RRA), a unit of S&P Global. The RRA database is comprehensive. It contains every electric utility rate case in the United States since 1980 in which the utility has requested a rate change of at least \$5 million or a regulator has authorized a rate change of at least \$3 million. Our study comprises the period from 1980 through 2018. Table 2 illustrates the bridge from the RRA rate-case population to the rate-case sample used in our analyses. We examined the returns on equity authorized by the regulatory agencies, *not* the returns requested by the utilities.[5] The sample we use in this paper contains 79% of the RRA universe, but 97% of the rate cases in which a rate of return on equity was authorized by a state regulator.

Nearly all fifty states and Washington D.C. are represented in the data set.[6] Thirty-two electric utility rate cases satisfying the qualifications listed above were filed in the average state over the past thirty-eight years, with the most being filed in Wisconsin (120) and the fewest being filed in Tennessee (3), Alaska (2), and Alabama (1). The frequency of filing in a state does not appear to have any relationship to premium growth. The average risk premium has grown in both the ten states that completed the most rate cases and the ten states that completed the fewest rate cases and has grown at very similar rates (see Fig. 1). In fact, as Fig. 2 illustrates, the general trend across all states is similar.

In the early 1980s there were over 100 rate cases filed each year. By the late 1990s, in the midst of widespread deregulation of the electric power industry, the number of filings reached its lowest point (with six in 1999). Since then, filing frequency has increased to an average of forty-eight per year over the last three years (see Fig. 3). The decline in rate case activity in many instances was the direct result of rate moratoria related to the transition to competitive markets in the late 1990s, as well as to moratorium-like concessions made to regulators related to merger approvals over the last decade. Many of these moratoria will expire over the next two years, suggesting a new increase in rate case activity is likely. Finally, no individual utility had an outsized influence on the sample. One hundred forty-four different companies filed rate cases, but many have since merged or otherwise stopped filing.[7] The average firm filed eleven rate cases in our sample. Within our sample the most frequently-filing entity was PacifiCorp, which filed seventy-three rate cases, or less than 5% of the sample.

3.2. Calculating the regulated equity premium

Regulated equity returns are generally equal to the sum of the riskless rate of return and a premium for risk-bearing. In the CAPM, the premium for risk-bearing is given by $\beta(r_m - r_f)$, where β is the utility's

[3] The other three approaches identified by Morin (2006) are: Risk Premium (which is an attempt to estimate empirically what the CAPM derives theoretically), Discounted Cash Flows (or "DCF," which is a dividend capitalization model), and Comparable Earnings (which is an empirical approach to deriving cost of capital from market comparables based on *Hope*).

[4] This has also been an ongoing issue of contention in recent regulatory proceedings. In Opinion 531-B (Federal Energy Regulatory Commission, March 3, 2015, 150 FERC 61,165), the Federal Energy Regulatory Commission (FERC) found that "anomalous capital market conditions" caused the traditional discount rate determination methods not to satisfy the *Hope* and *Bluefield* requirements (150 FERC 61,165 at 7). But in a related decision only eighteen months later (Federal Energy Regulatory Commission, September 20, 2016, 156 FERC 61,198), FERC acknowledged that expert witnesses disagreed as to whether any market conditions were, in fact, "anomalous" (156 FERC 61,198 at 10).

[5] To be clear, we refer to the rates set by regulators as the "authorized" rates. These may be contrasted with utilities' "requested" rates and also with the "earned" rates of return actually realized by utilities. Regulatory *authorization* of a rate is not a guarantee that a utility will actually *earn* such a rate. We address this issue in further detail in Section 4.5.

[6] Only Nebraska did not have a reported rate case meeting the parameters of the data set. Nebraska is unique in that it is the only state served entirely by consumer-owned entities (e.g., cooperatives, municipal power districts) and therefore absent a profit motive it does not have the same adversarial regulatory system as all other states.

[7] The level of analysis is at the regulated utility level. We recognize that many holding companies have multiple ring-fenced regulated utility subsidiaries.

Table 2
Bridge illustrating how our sample is constructed from the RRA electric utility rate case population data.

Number of cases	Percent of cases	Description
2033	100.0%	All electric utility rate cases 1980–2018 in which utility has requested a rate change of at least \$5 million or a regulator has authorized a rate change of at least \$3 million.
−19	−0.9%	Rate cases with final adjudication (i.e., fully-litigated or settled) still pending as of December 31, 2018, are excluded
−369	−18.2%	Rate cases with no return on equity determination are excluded
−30	−1.5%	Rate cases with no capital structure determination are excluded
−19	−0.9%	Rate cases with authorized rates lower than the then-prevailing riskless rate are excluded
1596	79.0%	Rate cases used in our analysis

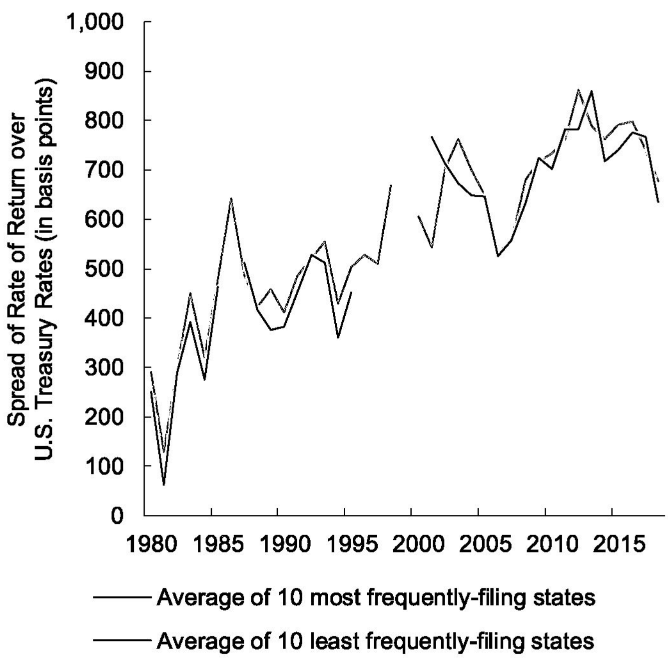


Fig. 1. Risk-premium growth by frequency of case filing. Gaps in the series reflect years in which no rate cases were filed for the subject group. The risk premium is calculated as $r_E - r_f$, or the excess of the authorized return on equity over the then-current riskless rate.

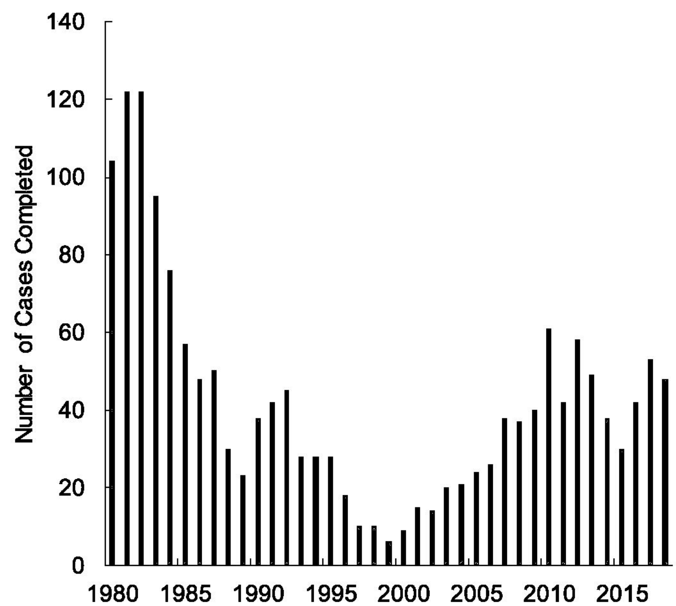


Fig. 3. Number of electric utility rate cases finalized by year.

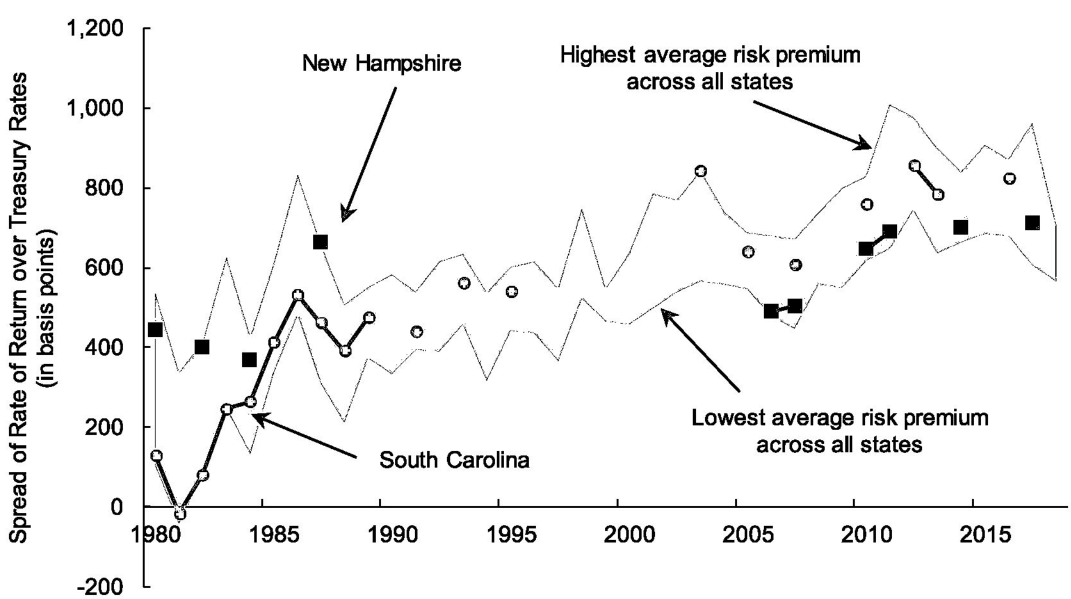


Fig. 2. Range of risk-premium growth across states. States with highest (New Hampshire) and lowest (South Carolina) rates of risk-premium growth over the period (among states with at least five rate cases) are highlighted.

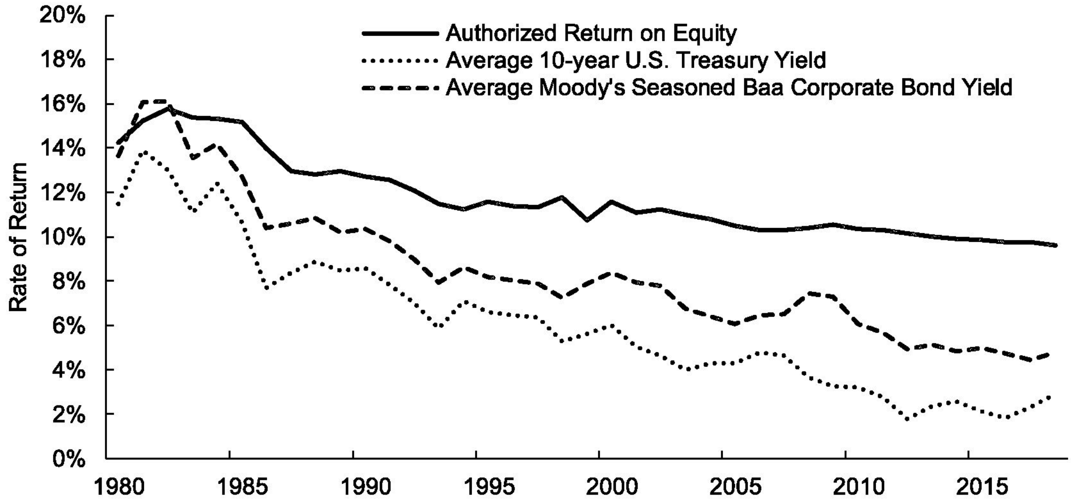


Fig. 4. Annual average authorized return on equity vs. U.S. Treasury and investment grade corporate bond rates.

equity beta. Rearranging the security market line equation [1], we define the regulated equity premium as $r_E - r_f = \beta(r_m - r_f)$. Presented thus, we first note that the existence of a (positive) regulated equity premium is not, by itself, evidence of irrational investor behavior or model failure. Neither is the existence of a growing regulated equity premium. We take no position here on what the "correct" premium should be in any instance. Rather, we shall be content in this article simply to determine whether or not the behavior of the risk premium in practice is consistent with financial theory.

On average, the authorized return on equity is 5.1% (standard deviation = 2.2%) higher than the riskless rate. Fig. 4 illustrates the average authorized return on equity over the period against the average annual riskless rate and investment-grade corporate bond rate.[8] For avoidance of doubt, we note that only the U.S. Treasury note rate should be considered the riskless rate. We include corporate bond rates solely to assess whether the trend in riskless rates is materially different from the trend in risky debt.

While the regulated equity premium has averaged 510 basis points across the entire time period, in 1980 the average premium was only 277 basis points, whereas in 2018 it averaged 668 basis points. Fig. 5 shows the difference between the authorized return on equity and the riskless rate for each case in the data over the past thirty-eight years. Although the premium is determined against the riskless rate of return (represented here as the yield on a 10-year U.S. Treasury note), we also present for comparison the spreads determined against the yield on the Moody's Seasoned Baa Corporate Bond Index to illustrate that the effect is not an artifact of recent monetary policy on Treasury rates. The trends of the two series are quite similar (and both have statistically-significant positive slopes); accordingly, we shall present only the Treasury rate-determined premiums throughout the remainder of this paper.

Given that a large and growing regulated equity premium exists, our question is whether or not it can be explained within an equilibrium asset-pricing framework such as the CAPM. If β were to have increased during the time period in question, for example, the growth of the regulated equity premium may well be explained by the increasing (relative) riskiness of utility equity. As Section 4 demonstrates, however, in fact it cannot.

[8] We used the 10-year constant maturity U.S. Treasury note yield as a proxy for the riskless rate and the yield on the Moody's Seasoned Baa Corporate Bond Index as a proxy for investment-grade corporate bond rates. Both series were obtained from the Federal Reserve's FRED database (Board of Governors of the Federal Reserve System, n.d.-a; n.d.-b).

4. Potential explanations for the premium

Having demonstrated the existence of a large and growing regulated equity premium, we investigate various potential explanations. As we indicated above, we proceed with our investigation of explanations for the premium via the Capital Asset Pricing Model. The CAPM allows three basic mechanisms of action for a change in the risk premium: (i) the manner in which the underlying assets are financed has changed, (ii) the risk of the underlying assets themselves has changed, and/or (iii) the rate at which the market in general prices risk has changed. We explore each in turn and formally test whether the trend in the data can be explained by the CAPM. Finding that it cannot, we then turn to theoretical explanations outside of the CAPM. The potential alternative explanations in Sections 4.5 through 4.7 all represent viable paths for further research.

4.1. *Capital structure effects*

As corporate leverage increases, the underlying equity becomes riskier and thus deserving of higher expected returns. In finance, the Hamada equation decomposes the CAPM equity beta (β) into an underlying asset beta (β_A) and the impact of capital structure (Hamada, 1969, 1972). Specifically, the Hamada equation states that $\beta = \beta_A\left[1 + (1-\tau)\frac{D}{E}\right]$, where τ is the tax rate and D and E are the debt and equity in the firm's capital structure, respectively. We use the marginal corporate federal income tax rate for the highest bracket, as provided in Internal Revenue Service (n.d.).

One explanation for a growing risk premium would be steadily increasing leverage among regulated utilities. However, regulators also generally approve of specific capital structures as part of the rate-making process. As a result, our database also contains the authorized capital structures for each utility.[9] In fact, utilities are *less* leveraged today than they were in 1980. The average debt-to-equity ratio in the first five years of the data set (1980–1984) was 1.74; in 2014–2018 it was 1.05. More generally, we can observe the impact of leverage

[9] To be clear, the authorized capital structures evaluated here apply to the regulated utility subsidiaries, and not necessarily to any holding companies to which they belong. The holding companies themselves may utilize more or less leverage, but typically the regulated utility subsidiaries are "ring-fenced" so as to isolate them from holding company-level risks. Similarly, rate-of-return regulation would apply only to the regulated subsidiaries, not to the parent holding company. As a result, the capitalization of the regulated entity (studied here) is often different from the capitalization of the publicly-traded entity that owns it.

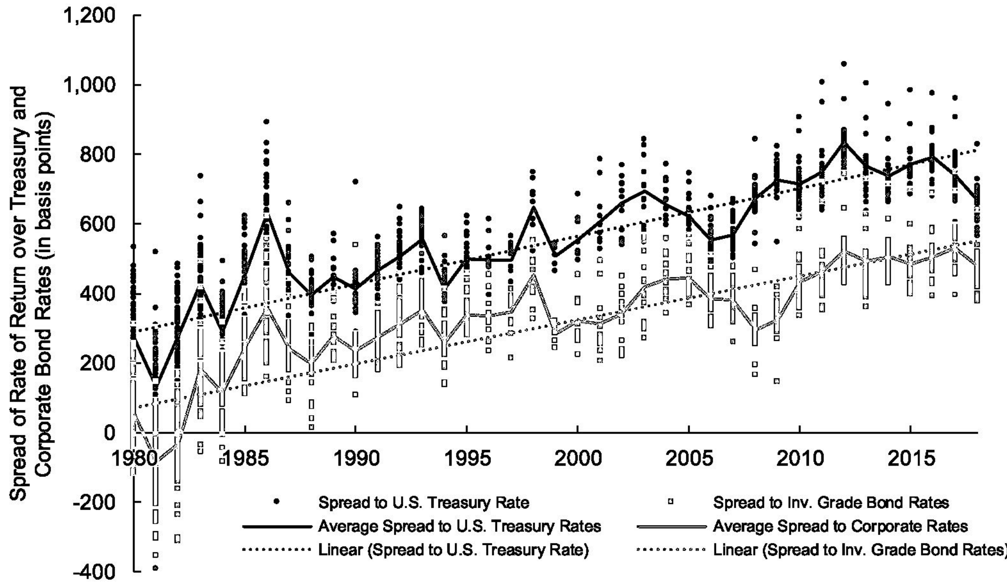


Fig. 5. Authorized return on equity premium, 1980–2018.

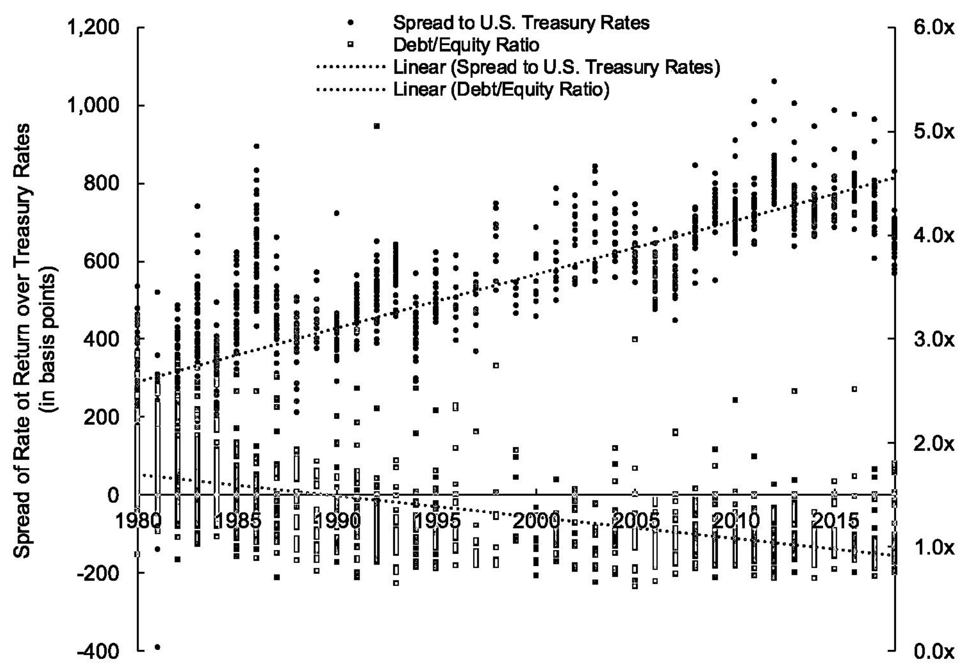


Fig. 6. Authorized return on equity premium vs. utility leverage.

moving in the opposite direction of what one may expect, whether we examine the debt-to-equity ratio exclusively or the Hamada capital structure parameter (i.e., the portion of the Hamada equation multiplied by β_A, or $\left[1 + (1-\tau)\frac{D}{E}\right]$) in its entirety. Figs. 6 and 7 illustrate these results. As a result, it does not appear as if capital structure itself can explain the behavior of the risk premium.

4.2. Asset-specific risk

As noted above, the Hamada equation decomposes returns into compensation for bearing asset-specific risks and for bearing capital structure-specific risks. Even if a firm's capital structure remains unchanged, the riskiness of its underlying assets may change. This risk is represented by the unlevered asset beta, β_A. An increase in the asset beta applicable to such investments would, all else held equal, justify an increase in the risk premium.

To examine such a hypothesis, we used the fifteen members of the Dow Jones Utility Average between 1980 and 2018 as a proxy for "utility asset risk." We estimated five-year equity betas for each firm by regression of their monthly total returns against the total return on the S&P 500 index.[10] The equity betas calculated were then converted to

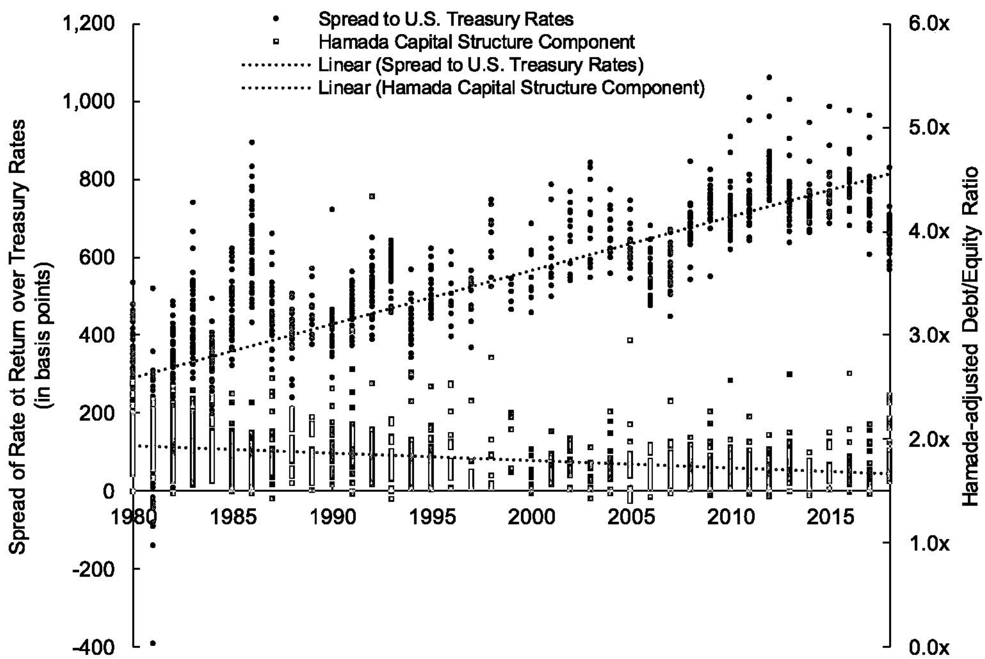


Fig. 7. Authorized return on equity premium vs. the Hamada capital structure parameter.

asset betas using Hamada's equation and corrected for firm cash holdings using firm-specific balance sheet information. We then averaged the fifteen asset betas calculated in each year as our proxy for utility asset risk.[11] The results remain substantively unchanged whether an equal-weighted or a capitalization-weighted average is used.

Although there is, of course, variation in the industry average asset beta across the thirty-eight years, the general trend is down. Fig. 8 presents the risk premium in comparison to the industry average asset beta. As a result, the asset beta is moving in the opposite direction from what one might expect, given a steadily-increasing risk premium, and therefore does not appear to explain the observed behavior of the risk premium.

4.3. The market risk premium

The last CAPM-derived explanation for a changing risk premium relates to the pricing of risk assets in general. If investors require greater compensation for bearing the systematic risk of the market in general, then the risk premium across all assets would increase as well (all else held equal) as a result of the average risk aversion coefficient of investors increasing. The market risk premium reflects this risk-bearing cost in the CAPM.

Although we can observe the *ex post* market risk premium, investors' assessment of the *ex ante* market risk premium is generally based on assuming that historical experience provides a meaningful guide to future experience.[12] It is customary to examine the actual market risk premium over some historical time period and base one's estimate of the *expected* future market risk premium on that historical experience (Sears and Trennepohl, 1993; Villadsen et al., 2017, p. 59). While the size of the historical window is subjective, it is sufficient for our purposes to note that the slope of the market risk premium over time has been negative irrespective of the historical window used.[13] Most sources advocate for using the longest time window available (Villadsen et al., 2017, p. 61); we use a fifty-year historical window for calculation purposes. As Fig. 9 illustrates, that declining trend in the market risk premium appears to be inconsistent with the increasing risk premium exhibited by the rates of return authorized by regulators.

4.4. Testing a theoretical model of the risk premium

Although we have illustrated that each component of the CAPM risk premium appears at odds with the risk premium derived from rates of return authorized by regulators, we now turn to a formal exploration of these relationships. By combining the security market line representation of the CAPM [1] and the Hamada equation, we can define the risk premium, $r_E - r_f$.

$$r_E - r_f = \beta_A \times \left[1 + (1-\tau)\frac{D}{E}\right] \times MRP \tag{2}$$

In [2], $r_E - r_f$ is the risk premium, or the difference between the authorized rate of return on equity for a given firm in a given rate case and the then-prevailing riskless rate. The asset beta, β_A, is calculated as described in Section 4.2. The middle component is taken from the Hamada equation and reflects the marginal corporate income tax rate (τ) in effect in the year in which the equity return was authorized and the authorized debt-to-equity ratio reflected in the regulators' decision for each case. Lastly, *MRP* is the *ex ante* estimate of the market risk

[10] We determined the composition of the Dow Jones Utility Average index at the end of each year and used a rolling five-year window to perform the regressions. For example, the 1980 regression betas were estimated based on monthly returns from 1975 to 1979, the 1981 regression betas were estimated based on monthly returns from 1976 to 1980, and so on.

[11] The balance sheet and total return data are taken from Standard & Poor's COMPUSTAT database. We calculate $\beta'_A = \beta / \left[1 + (1-\tau)\frac{D}{E}\right]$ and $\beta_A = \beta'_A / \left[1 - \frac{C}{D+E}\right]$, where C equals the amount of cash and cash equivalents held by each firm and D and E represent, respectively, the debt and equity of each firm. We measure D as the sum of Current Liabilities, Long-Term Debt, and Liabilities–Other in the COMPUSTAT data. Because final firm accounting information was not available for 2018 at the time of writing, we maintained the capital structures calculated using 2017 data.

[12] We do not dwell here on the issue of the "observability" of the market portfolio as it relates to testability of the CAPM. We shall assume that the S&P 500 index is a reasonable proxy for the market portfolio.

[13] The market risk premium data used here are taken from data on the S&P 500 and 10-year U.S. Treasury notes collected from the Federal Reserve (Damodaran, n.d.).

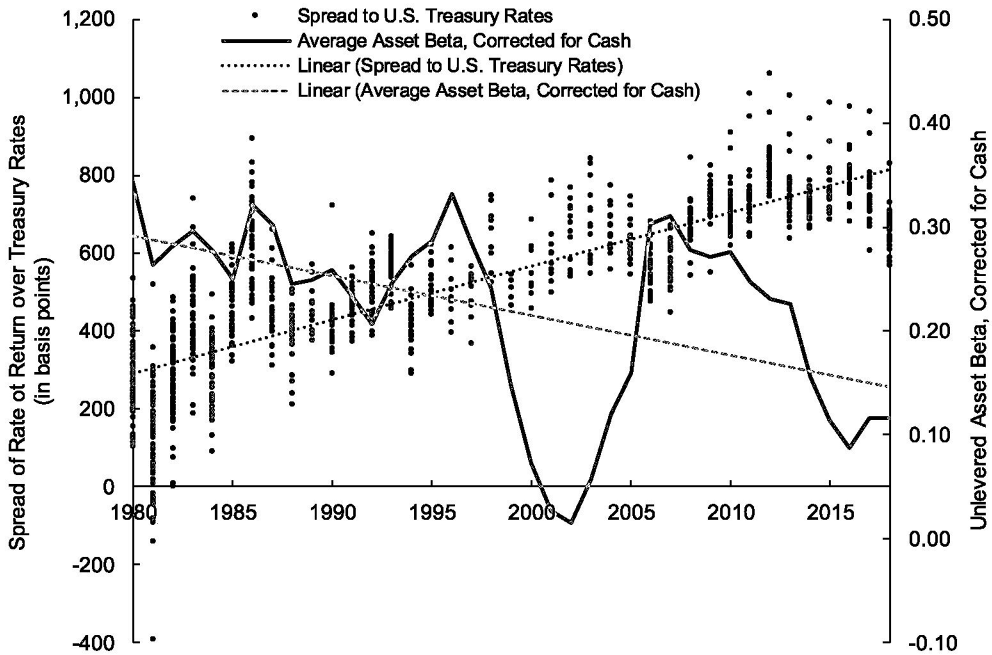


Fig. 8. Authorized return on equity premium vs. industry average asset beta.

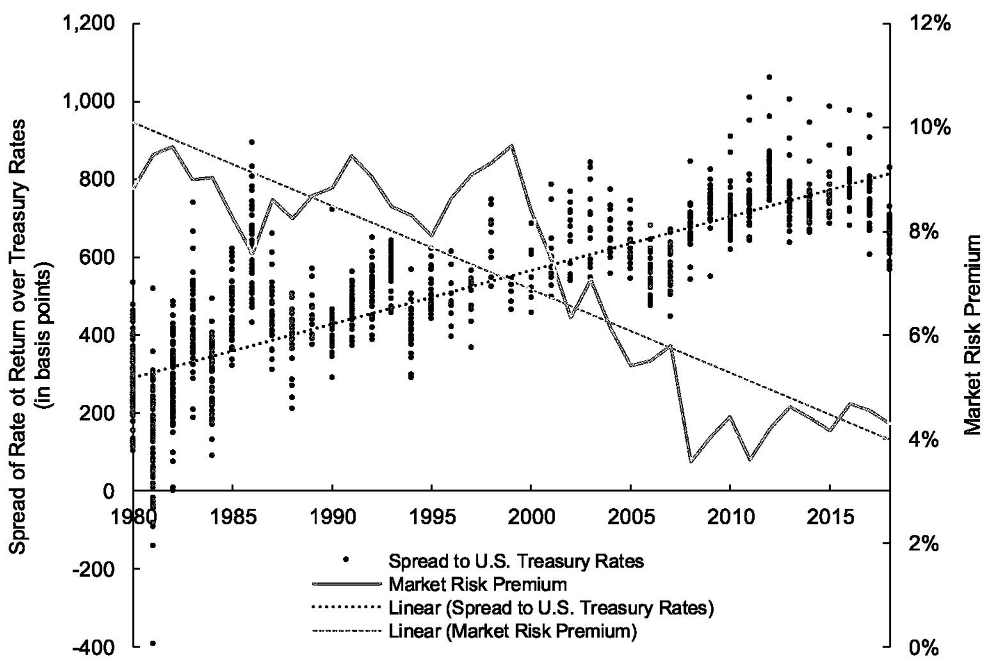


Fig. 9. Authorized rate-of-return premium vs. *ex ante* estimated market risk premium.

premium based on a fifty-year historical window as of the year in which each equity return was authorized.

Let $i = 1, ..., N$ index firms and $t = 1, ..., T$ index years. Not every firm files a rate case in every year. In addition, firms enter and exit over time due to merger or bankruptcy. Because regulators must have an evidentiary record to support their determinations, we assume that each rate case is evaluated independently in an adversarial hearing across time.

By using a logarithmic transform of [2], we arrive at equation [3].

$$ln(r_{E,it} - r_{f,t}) = \gamma_0 + \gamma_1 ln(\beta_{A,t}) + \gamma_2 ln\left[1 + (1 - \tau_t)\frac{D_{it}}{E_{it}}\right] + \gamma_3 ln(MRP_t) \tag{3}$$

In a traditional ordinary least squares (OLS) regression setting, the CAPM would hypothesize that γ_0 should be zero (or not significant) and γ_1, γ_2, and γ_3 should be positive and significant. What we find, however, is exactly the opposite of that (Table 3). The coefficients are negative and strongly significant. Further, a comparison of the observed risk premium to the risk premium estimated by our regression model reveals a good fit (Fig. 10). The negative coefficients are problematic for the CAPM, but also suggest rather counterintuitive effects at an applied

Table 3
Regression results for CAPM-based risk premium model. Coefficients for both the OLS regression model and a model controlling for utility-level fixed effects are shown.

	OLS	Controlling for utility-level fixed effects
	$ln(r_E - r_f)$	$ln(r_E - r_f)$
γ_0, Constant	3.641****	
	(0.130)	
γ_1, Asset beta, $ln(\beta_A)$	−0.158****	−0.156****
	(0.022)	(0.023)
γ_2, Capital structure, $ln\left[1 + (1-\tau)\frac{D}{E}\right]$	−0.492****	−0.967****
	(0.103)	(0.142)
γ_3, Market risk premium, $ln(MRP)$	−0.947****	−0.898****
	(0.035)	(0.039)
R-squared	46.4%	46.6%
Adjusted R-squared	46.3%	41.2%
F statistic	458.8****	420.9****
No. of observations	1596	1596

Standard errors are reported in parentheses.
*, **, ***, and **** indicate significance at the 90%, 95%, 99%, and 99.9% levels, respectively.

level. Regulators use CAPM prescriptively in rate cases; they are determining what utilities *should* earn. A negative capital structure coefficient suggests, for example, that investors in firms with high leverage *should* be compensated with *lower* returns. Similarly, negative coefficients imply that investors in firms with riskier assets (higher asset betas) and during periods of higher risk aversion (higher market risk premiums) should also be compensated with *lower* returns. These results would be difficult for regulators to justify on normative grounds.

It may be the case, however, that common cross-sectional variation is biasing the results for this data by creating endogeneity issues for the OLS-estimated coefficients. For example, the repeated presence of the same utilities over time could introduce entity-level fixed effects into the analysis. Accordingly, we performed an F-test to evaluate the presence of individual-level effects in the data (Judge et al., 1985: p. 521). The test strongly supports the presence of individual (utility-level) effects ($F_{143,1449} = 1.5$, $p < 0.001$). In addition, the Hausman test (Hausman, 1978; Hausman and Taylor, 1981) supports the fixed-effect specification in lieu of random effects ($\chi^2(3) = 24.0$, $p < 0.001$). As a result, Table 3 also provides the regression coefficients controlling for utility-level fixed effects. These coefficients, while numerically different than the OLS results, are nevertheless still negative and strongly significant, in conflict with both financial theory and regulator intent.

Fig. 10 also reveals a distinct shift in the predicted trend of the risk premium beginning in 1999. This is notable because for many parts of the U.S., 1999 represented the year that implementation of electric market reform and restructuring began, with wholesale markets such as ISO-New England opening and several divestiture transactions of formerly-regulated generating assets occurring, establishing market valuations for formerly regulated assets (Borenstein and Bushnell, 2015). In addition, FERC issued its landmark Order 2000 encouraging the creation of Regional Transmission Organizations. To examine this point in time, we divided the data into two sets, 1980–1998 and 1999–2018, and estimated separate regression models for each subset using both OLS and controlling for utility-level fixed effects (Table 4). As before, the F (pre-1999 $F_{129,805} = 1.6$, $p < 0.001$; post-1998 $F_{129,525} = 3.2$, $p < 0.001$) and Hausman (pre-1999 $\chi^2(3) = 15.5$, $p < 0.01$; post-1998 $\chi^2(3) = 23.8$, $p < 0.001$) tests both strongly support the model controlling for utility-level fixed effects over OLS.

Although the results in both cases are consistent with our earlier finding that the standard finance model appears at odds with the empirical data, the two regression models are noticeably different from one another and appear to better represent the data (Fig. 11). We performed the Chow (1960) test and confirmed the presence of a structural break in the data in 1999 ($F_{4,1588} = 91.6$, $p < 0.001$).[14] We find this result suggestive that deregulatory activity may have an influence even on still-regulated utilities—a point to which we shall return in Section 5.2.

4.5. Potential finance explanations other than the CAPM

In Mehra and Prescott's (2003) review of the equity premium puzzle literature, the authors acknowledge that uncertainty about changes in the prevailing tax and regulatory regimes may explain the premium. Such forces may also be at work with regard to regulated rates of return. To the extent that investors require higher current rates of return because they are concerned about future shocks to the tax or regulatory structure of investments in regulated electric utilities (e.g., EPA's promulgation of the Clean Power Plan, the U.S. Supreme Court's stay of the Clean Power Plan, expiration of tax credits), such concern may be manifest in a higher degree of risk aversion that is unique to investors in the electric utility sector, and therefore a higher "market" risk premium on the assumption that capital markets are segmented for electric utilities.

A separate line of inquiry concerns a criticism of the Hamada equation in the presence of risky debt (Hamada (1972) excluded default from consideration). Conine (1980) extended the Hamada equation to accommodate risky debt by applying a debt beta. Subsequently, Cohen (2008) sought to extend the Hamada equation by adjusting the debt-to-equity parameter to incorporate risky debt in the calculation of the equity beta [4].

$$\beta = \beta_A\left[1 + (1-\tau)\frac{r_D}{r_f}\frac{D}{E}\right] \tag{4}$$

We view neither of these proposed solutions as entirely satisfying, and note that they tend to be material only for high leverage, which is not common to regulated utilities. Nevertheless, we acknowledge that adjustments to the capital structure may influence the risk premium. However, applying the Cohen (2008) modification and using the Moody's Seasoned Baa Corporate Bond Yield as a proxy for the cost of risky debt (r_D), we note that our regression results are substantively unchanged. As Table 5 illustrates, use of the Cohen betas still results in highly significant, but negative coefficients, which is contrary to theory. These results are maintained when controlling for utility-level fixed effects, and the F (Hamada $F_{143,1449} = 1.5$, $p < 0.001$; Cohen $F_{143,1449} = 1.3$, $p < 0.01$) and Hausman (Hamada $\chi^2(3) = 24.0$, $p < 0.001$; Cohen $\chi^2(3) = 6.3$, $p < 0.1$) tests are significant in support of the fixed effects model.

In lieu of modifying the CAPM parameters, some researchers have suggested that Ross's (1976) Arbitrage Pricing Theory (APT) is preferable to the CAPM because the CAPM produces a "shortfall" in estimated returns (Roll and Ross, 1983) and "underestimates" actual returns in utility settings (Pettway and Jordan, 1987). While the works of these authors are suggestively similar to the analysis contained in this paper, we note that those authors were examining the *actual* returns on utility common stocks, rather than the rates of return *authorized* by regulators for assets held in utility rate bases. The distinction is important. In the case of the former, it is a question of asset pricing models and efficient capital markets. In the case of the latter, it is an issue of regulator judgment. We note specifically that regulators are making decisions that set these rates, and in many cases are doing so explicitly stating that they are relying in whole or in part on the CAPM. Our question concerns not just whether the CAPM is a better asset pricing model (than the APT, for example), but whether regulators' own judgment can

[14] Additional testing using the Andrews (1993) approach supports the presence of structural breaks during the transitional regulatory period identified by Borenstein and Bushnell (2015), confirming the appropriateness of our selection of 1999 as a year with strong historical motivation for a structural break.

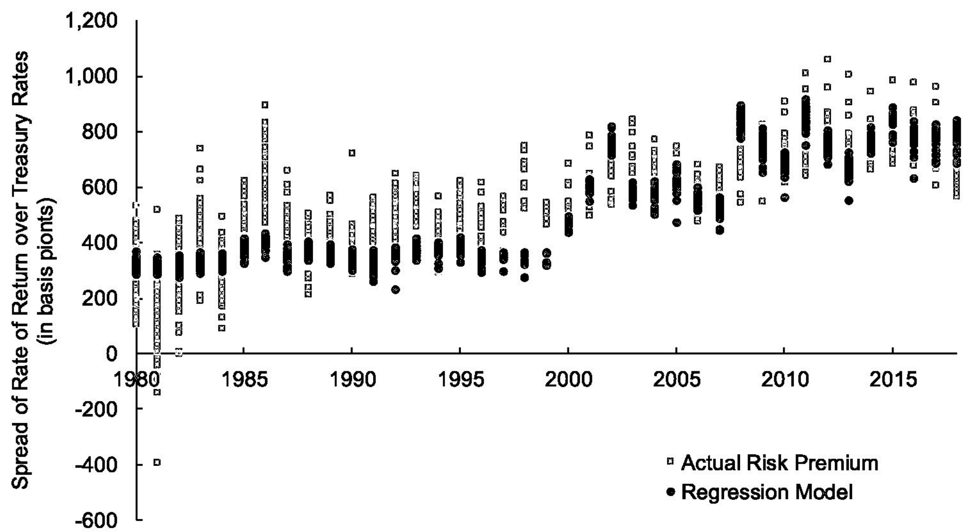


Fig. 10. Actual vs. OLS regression-model risk premium.

Table 4
Regression results for a two-period CAPM-based risk premium model. For purposes of the Chow test, the combined sum of squared residuals was 272.5. Coefficients for both the OLS regression model and a model controlling for utility-level fixed effects are shown.

	OLS		Controlling for utility-level fixed effects	
	1980–1998	1999–2018	1980–1998	1999–2018
	$ln(r_E - r_f)$	$ln(r_E - r_f)$	$ln(r_E - r_f)$	$ln(r_E - r_f)$
γ_0, Constant	−6.259****	5.159****		
	(0.718)	(0.093)		
γ_1, Asset beta, $ln(\beta_A)$	−0.940****	−0.071****	−0.972****	−0.065****
	(0.131)	(0.008)	(0.135)	(0.008)
γ_2, Capital structure, $ln\left[1 + (1 - \tau)\frac{D}{E}\right]$	−0.140	−0.325****	−0.865****	−0.636****
	(0.150)	(0.049)	(0.224)	(0.075)
γ_3, Market risk premium, $ln(MRP)$	−4.529****	−0.471****	−4.326****	−0.432****
	(0.261)	(0.026)	(0.267)	(0.025)
R-squared	26.7%	36.9%	30.2%	44.9%
Adjusted R-squared	26.4%	36.6%	18.8%	31.0%
F statistic	113.3****	127.3****	116.0****	142.5****
Sum of squared residuals	214.4	8.4	170.8	4.7
No. of observations	938	658	938	658

Standard errors are reported in parentheses.
*, **, ***, and **** indicate significance at the 90%, 95%, 99%, and 99.9% levels, respectively.

be explained by the model on which they claim to rely.

Lastly, to address a related point, we also examined the actual earned rates of return on equity for the 15 utilities in the Dow Jones Utility Average over our historical window. We used each firm's actual return on equity, calculated annually as Net Income divided by Total Equity, as reported in the COMPUSTAT database. This measure of firm profitability examines how successful the firms were at converting their *authorized* returns into *earned* returns. In general, the earned returns closely tracked the authorized returns, suggesting that the decisions of regulators are significantly influencing the actual earnings of regulated utilities. Fig. 12 compares the spread of *authorized* rates of return over riskless rates to the spread of *earned* rates of return over riskless rates and to the median net income of utilities in constant 2018 dollars.[15] The steadily increasing risk premium we have identified is present in both series. The series are correlated at 0.77 (authorized vs. earned), 0.59 (authorized vs. median net income), and 0.75 (earned vs. median net income), all of which are significantly greater than zero ($p < 0.001$). Further, the "capture rate" (the percentage of authorized rates actually earned by the utilities) averaged 96% over the entire time period. As a result, we conclude that the trend of increasing risk premiums is not an abstract anomaly occurring in a regulatory vacuum, but rather a direct contributor to the earnings of regulated utilities.

However, these measures of firm performance must be interpreted with caution. The authorized rates of return apply to jurisdictional utilities, while the earned rates of return are calculated based on holding company performance, which in many cases are not strictly equivalent. Further, increasing net income may be due to industry consolidation producing larger firms (with income increasing only proportionally to size), rather than an increase in profitability itself. In fact, the average income-to-sales ratio of the Dow Jones Utility Average members remained remarkably stable across the period of our study,

[15] We used the median earned rate of return over the 15 Dow Jones utilities. The results are substantively equivalent if the average earned rate of return is used but are more volatile due to the impact on earnings of the California energy crisis of 2000–2001 and the collapse of Enron in 2001.

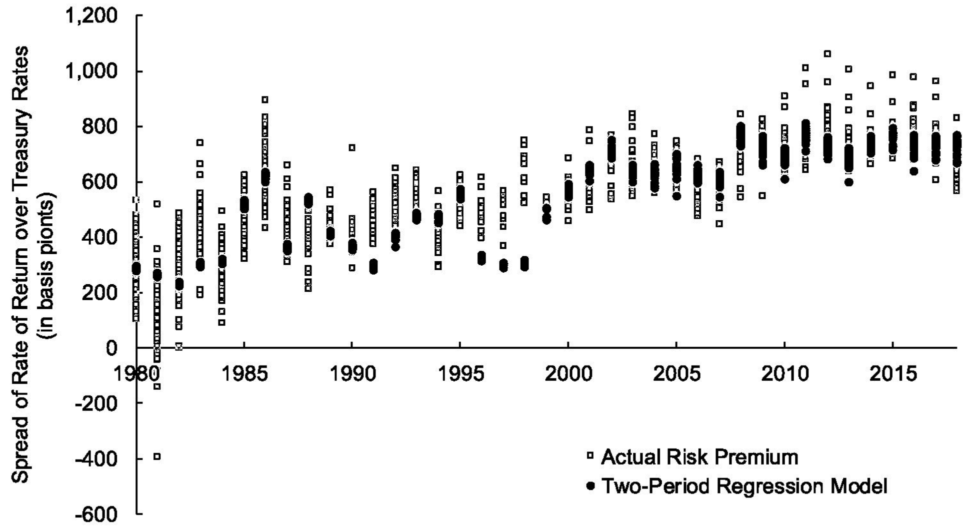


Fig. 11. Actual vs. two-period OLS model-predicted risk premium.

Table 5
Regression results for the standard Hamada capital structure model and Cohen (2008) capital structure model that incorporates risky debt. Coefficients for both the OLS regression model and a model controlling for utility-level fixed effects are shown.

	OLS		Controlling for utility-level fixed effects	
	Hamada $ln(r_E - r_f)$	Cohen $ln(r_E - r_f)$	Hamada $ln(r_E - r_f)$	Cohen $ln(r_E - r_f)$
γ_0, Constant	3.641****	3.191****		
	(0.130)	(0.085)		
γ_1, Asset beta, $ln(\beta_A)$	−0.158****	−0.169****	−0.156****	−0.175****
	(0.022)	(0.022)	(0.023)	(0.023)
γ_2, Capital structure, $ln\left[1 + (1 - \tau)\frac{D}{E}\right]$	−0.492****		−0.967****	
	(0.103)		(0.142)	
γ_2', Capital structure, $ln\left[1 + (1 - \tau)\frac{r_D}{r_f}\frac{D}{E}\right]$		−0.156*		−0.275***
		(0.081)		(0.040)
γ_3, Market risk premium, $ln(MRP)$	−0.947****	−1.046****	−0.898****	−1.087****
	(0.035)	(0.036)	(0.039)	(0.040)
R-squared	46.4%	45.7%	46.6%	45.1%
Adjusted R-squared	46.3%	45.6%	41.2%	39.6%
F statistic	458.8****	447.1****	420.9****	396.9****
No. of observations	1596	1596	1596	1596

Standard errors are reported in parentheses.
*, **, ***, and **** indicate significance at the 90%, 95%, 99%, and 99.9% levels, respectively.

and actually slightly declined, suggesting that gains in net income came from growing revenue, rather than increasing margins (although revenue growth may itself be a function of rising authorized rates of return). Nevertheless, the results are suggestive.

We have not repeated the analysis of Roll and Ross (1983) and Pettway and Jordan (1987) and examined the relationship between firm performance and stock performance. Their findings, however, suggest that regulated utilities have realized *higher* stock returns than can be explained by the CAPM—a finding congruent with our work and suggestive of other factors being priced by the market. This does not entirely explain the judgment issue, however: why regulators appearing to use a CAPM approach provide utilities with returns that also appear to be excessive.

4.6. Potential public choice explanations

Another category of potential explanations emerges from the public choice literature on the role of institutional factors. Regulators may be deliberately or inadvertently providing a "windfall" of sorts to electric utilities. Stigler (1971), among others in the literature on regulatory capture, noted that firms may seek out regulation as a means of protection and self-benefit. This is particularly true when the circumstances are present for a collective action problem (Olson, 1965) of concentrated benefits (excess profits to utilities may be significant) and diffuse costs (the impact of those excess profits on each individual ratepayer may be small). Close relationships between regulators and the industries that they regulate have been observed repeatedly, and one possible explanation for the size and growth of the risk premium is the electric utility industry's increasing "capture" of regulatory power.

We are somewhat skeptical of this explanation, however, both because of the degree of intervention in most utility rate cases by non-utility parties, and because the data do not suggest that regulators have become progressively laxer over time. Fig. 13 compares the rates of return on equity *requested* by utilities in our data set against the rates of return ultimately authorized. As the trend line illustrates, this ratio has remained remarkably stable (within a few percent) over the thirty-eight

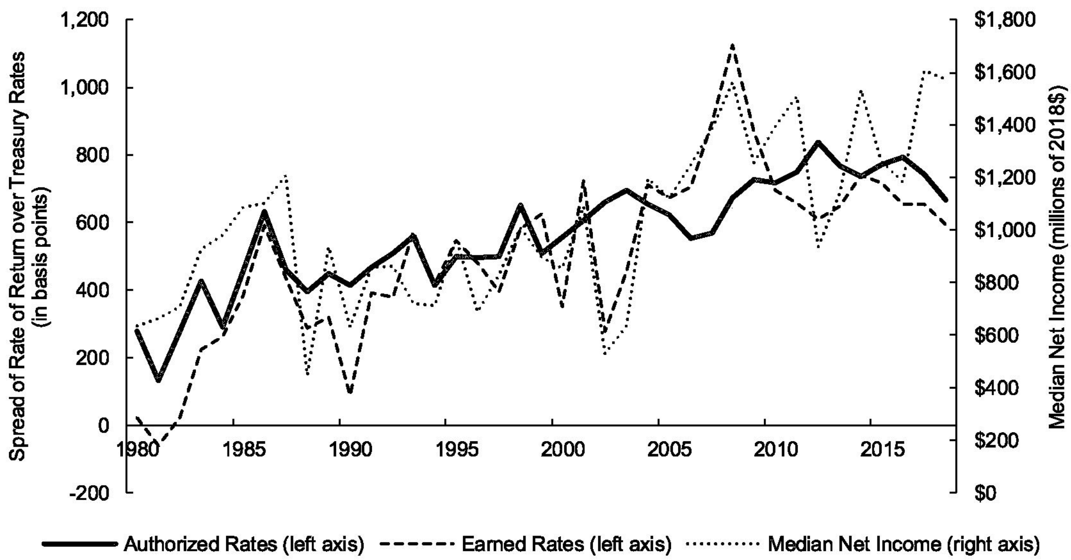


Fig. 12. Comparability of spreads measured with authorized and earned rates of return and utility net income.

years of data, even as the risk premium itself has steadily increased. As a result, the data do not suggest in general an obvious, growing permissiveness on the part of regulators. However, the last nine years are *suggestive* of an increasing level of accommodation among regulators. We propose a possible explanation for this particular pattern in Section 4.7.

To examine the public choice issues further, we investigated whether the risk premiums were related to the selection method of public utility commissioners and whether or not the rate cases in question were settled or fully litigated. The traditional hypothesis has been that elected (instead of appointed) commissioners were less susceptible to capture, more "responsive" to the public, and therefore more pro-consumer. Further, that cases that were settled were more likely to be accommodating to utilities (as money was "left on the table") and therefore would result in higher rates.

A sizable body of literature, however, has largely rejected the selection method hypothesis. Hagerman and Ratchford (1978) and Primeaux and Mann (1986) concluded that the selection method had no impact on returns or electricity prices respectively. Others have agreed that the selection method alone doesn't matter; it is how closely the regulators selected are monitored that matters (Boyes and McDowell, 1989). In addition, whatever evidence of an effect that may exist is likely due to selection method being a proxy for states with different intrinsic structural conditions (Harris and Navarro, 1983). Lastly, while states with elected utility commissioners (Kwoka, 2002) or commissioners whose appointment by the executive requires approval by the legislature (Boyes and McDowell, 1989) tend to have lower electricity prices, those low prices may create the perception of an "unfavorable" investment climate and may therefore lead to a higher cost of capital (Navarro, 1982). Alternatively, if lower prices are observed, it then remains unclear who actually pays (utility shareholders in foregone profits or consumers in higher costs of capital) for the lower observed prices (Besley and Coate, 2003).

To examine the impact of commission selection method and means of case resolution on risk premium, we categorized each state as having an elected or appointed utility commission based on data in Costello (1984), Besley and Coate (2003), and Advanced Energy Economy (2018). In addition, each rate case was reported as being either fully litigated or settled. The literature has hypothesized (but largely not found) that elected commissions are more "responsive" and therefore more pro-consumer. As a result, the expectation would be that the risk premiums implicit in authorized rates were higher for appointed commissions. Similarly, for means of case resolution, risk premiums would

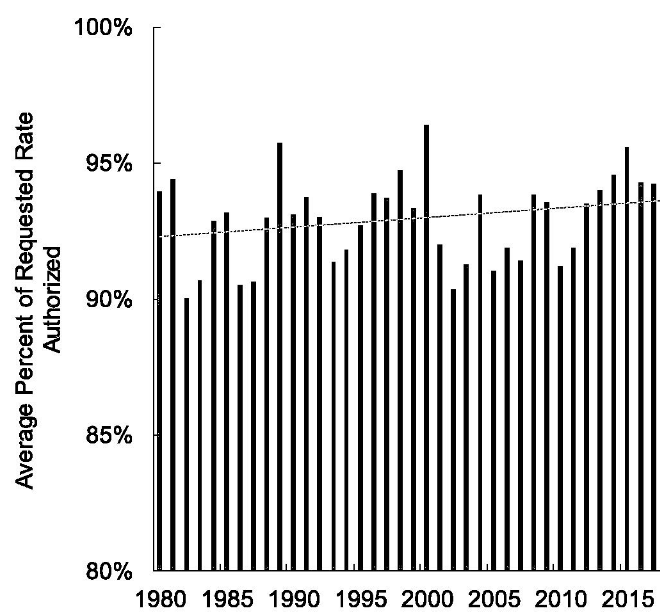


Fig. 13. Rate of return authorized as a percent of rate of return requested.

Table 6
Average risk premium in basis points by commission selection method and means of case resolution. The number of cases in each group is provided in parentheses.

	Appointed Commissions	Elected Commissions	Subtotals
Settled Cases	612 (367)	697 (89)	629 (456)
Fully Litigated Cases	460 (1008)	488 (181)	464 (1189)
Subtotals	500 (1375)	557 (270)	510 (1645)

be higher for settled, rather than fully litigated rate cases.

Like other authors, we found no significant effect *overall* for selection method, but a very significant effect for whether cases were settled or fully litigated. In addition, there appears to be a significant *interaction* between selection method and means of case resolution, suggesting that the lack of evidence of an effect in the literature may be related to its interaction with the means of case resolution, which has not been examined in this depth before. Table 6 illustrates the average risk

Table 7
Regression results for the standard CAPM model and the CAPM model plus two public choice variables (commission selection method and means of case resolution). Coefficients for both the OLS regression model and a model controlling for utility-level fixed effects are shown.

	OLS		Controlling for utility-level fixed effects	
	CAPM $ln(r_E - r_f)$	CAPM + Public Choice $ln(r_E - r_f)$	CAPM $ln(r_E - r_f)$	CAPM + Public Choice $ln(r_E - r_f)$
γ_0, Constant	3.641****	3.519****		
	(0.130)	(0.137)		
γ_1, Asset beta, $ln(\beta_A)$	−0.158****	−0.159****	−0.156****	−0.154****
	(0.022)	(0.022)	(0.023)	(0.023)
γ_2, Capital structure, $ln\left[1 + (1-\tau)\frac{D}{E}\right]$	−0.492****	−0.463****	−0.967****	−0.917****
	(0.103)	(0.102)	(0.142)	(0.141)
γ_3, Market risk premium, $ln(MRP)$	−0.947****	−0.927****	−0.898****	−0.858****
	(0.035)	(0.036)	(0.039)	(0.041)
γ_4, Settle = 1		0.223***		0.249****
		(0.057)		(0.060)
γ_5, Appointed = 1		0.159****		0.132**
		(0.034)		(0.058)
γ_6, Settle = 1 × Appointed = 1		−0.182***		−0.197***
		(-0.061)		(-0.065)
R-squared	46.4%	47.4%	46.6%	47.3%
Adjusted R-squared	46.3%	47.2%	41.2%	41.9%
F statistic	458.8****	238.5****	420.9****	216.5****
AIC	−2809	−2810		
No. of observations	1596	1596	1596	1596

Standard errors are reported in parentheses.
*, **, ***, and **** indicate significance at the 90%, 95%, 99%, and 99.9% levels, respectively.

premium observed in each group. The average risk premium for settled cases is significantly higher than for fully litigated cases ($p < 0.001$). Further, while the average risk premium for settled cases and appointed commissions is significantly greater than for fully litigated cases and elected commissions ($p < 0.001$), there is an interaction effect suggesting that the impact of selection method on risk premium depends on the means of case resolution ($p < 0.05$).

Notwithstanding these differences, the incremental explanatory value of these public choice variables is minimal (but significant). Table 7 compares the standard CAPM model with an OLS model that incorporates selection method and means of case resolution. The Akaike Information Criterion (AIC) indicates that incorporation of the public choice variables has only slight incremental value. We estimate that the marginal impact is only approximately 8 basis points—much less than the observed increase over time.[16] As before, the F (CAPM $F_{143,1449} = 1.5$, $p < 0.001$; CAPM + Public Choice $F_{143,1446} = 1.4$, $p < 0.001$) and Hausman (CAPM $\chi^2(3) = 24.0$, $p < 0.001$; CAPM + Public Choice $\chi^2(6) = 24.1$, $p < 0.001$) tests strongly support controlling for utility-level fixed effects in the model. Table 7 also includes coefficients incorporating such controls.

4.7. Potential behavioral economics explanations

To this point, we have examined a number of factors related to economic and institutional influences. At the outset, however, we noted the potential for rate determination to be influenced by regulator judgment. In many cases there is evidence that regulators are not behaving in accordance with the method they in fact purport to be using (i.e., CAPM). As we cannot escape the fact that ultimately the authorized return on equity is a product of regulator decision-making, we now consider possible explanations for the risk premium based on insights from behavioral economics.

First, we note that regulator attachment to rate decisions from the recent past may be coloring their forward-looking decisions. Earlier we referenced a report from Pennsylvania regulators about their stated reliance on (*inter alia*) "recent [returns on equity] adjudicated by the Commission" (Pennsylvania Public Utility Commission, 2016, p. 17). The legal weight attached to precedent may give rise here to a recency bias, where regulators anchor on recent rate decisions and insufficiently adjust them for new information. While stability in regulatory decision-making is seen as useful in assuring investors, to the extent that it results in a slowing of regulatory response when market conditions change, regulators should be encouraged to weigh the benefits of stability against the costs of distortionary responses to authorized returns that lag market conditions.

Our second insight from behavioral economics involves a curious observation in the empirical data: the average rate of return on regulated equity appears to have "converged" to 10% over time. Although the underlying riskless rate has continued to drop, authorized equity returns have generally remained fixed in the neighborhood of 10%, only dropping below (on average) over the last few years. Anecdotally, we have observed a reluctance among potential electric power investors to accept equity returns on power investments of less than 10%—even though those same investors readily acknowledge that *debt* costs have fallen. To that extent, then, a behavioral bias may be at work.

The finance literature has noted a similar effect related to crossing index threshold points (e.g., every thousand points for the Dow Jones Industrial Average). These focal points, which have no normative import, appear to influence investor behavior. Trading is reduced near major crossings (Donaldson and Kim, 1993; Koedijk and Stork, 1994; Aragon and Dieckmann, 2011), with some asserting that the behavior of investors in clienteles may produce this behavior (Balduzzi et al., 1997). We propose a related theory.

In economics, "money illusion" refers to the misperception of nominal price changes as real price changes (Fisher, 1928). Shafir et al. (1997) proposed that this type of choice anomaly arises from framing effects, in that individuals give improper influence to the nominal representation of a choice due to the convenience and salience of the nominal representation. The experimental results have been upheld in several subsequent studies in the behavioral economics literature (Fehr and Tyran, 2001; Svedsäter et al., 2007).

The effect here may be similar: investors and regulators may conflate "nominal" rates of return (the authorized rates) with the risk

[16] For example, the marginal impact of a settled vs. fully-litigated case would be $exp(3.513 + 0.223) - exp(3.513) = 8.4$ using the OLS coefficients.

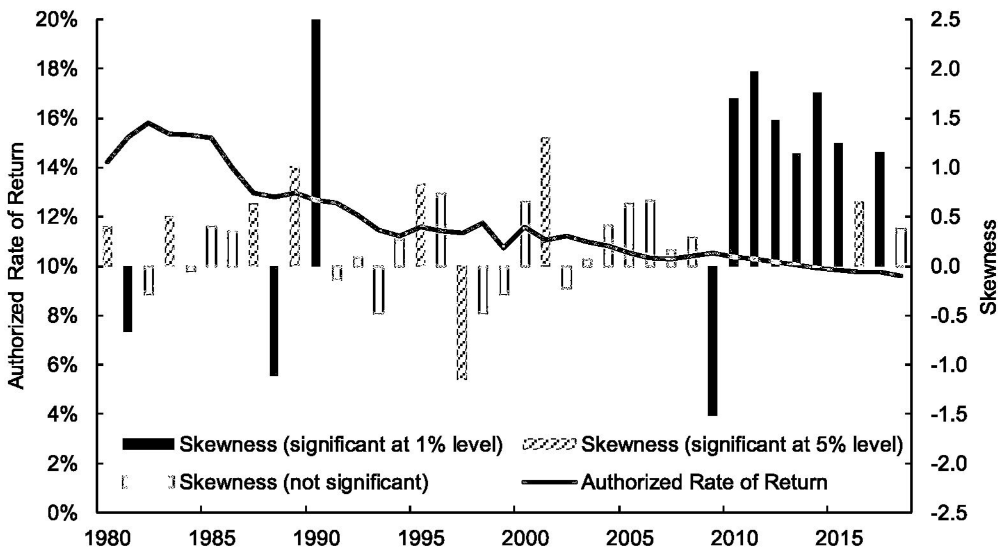


Fig. 14. Authorized rates of return on equity and skewness.

premium underlying the authorized rate. The apparent "stickiness" of rates of return on equity around 10% is similar to the "price stickiness" common in the money illusion (and, indeed, the rate of return is the price of capital). If there was in fact a tendency (intentional or otherwise) to respect a 10% "floor," one might expect that the distribution of authorized returns within each year may "bunch up" in the left tail at 10%, where absent such a floor one may expect them to be distributed symmetrically around a mean. As Fig. 14 illustrates, we see precisely such behavior. As average authorized returns decline to 10% (between 2010 and 2015), the skewness of the within-year distributions of returns becomes persistently and statistically significantly positive, suggesting a longer right-hand tail to the distributions, consistent with a lack of symmetry below the 10% threshold.[17] We note also that this period of statistically significant positive skewness coincides precisely with what appeared to be a period of increased regulator accommodation in Fig. 13. Further, once the threshold is definitively crossed, skewness appears to moderate and the distribution of returns appears to revert toward symmetry.

A related finding has been reported by Fernandez and colleagues (Fernandez et al., 2015, 2017, 2018), where respondents to a large survey of finance and economics professors, analysts, and corporate managers tended, on average, to overestimate the riskless rate of return. In addition, their estimates exhibited substantial positive skew, in that overestimates of the riskless rate far exceed underestimates.[18] The authors found similar results not just in the U.S., but also in Germany, Spain, and the U.K. In the U.S., the average response during the high skewness period exceeded the contemporaneous 10-year U.S. Treasury rate by 20–40 basis points, before declining as skewness moderated in 2018. It may be that overestimating the riskless rate is simply one way for participants in regulatory proceedings to "rationalize" maintaining the authorized return in excess of 10%. Alternatively, it may be an additional bias in the determination of authorized rates of return.

If such biases exist, there are clear implications for the regulatory function itself. For example, this apparent 10% "floor" was even recognized recently in a U.S. Federal Energy Regulatory Commission proceeding (Initial Decision, Martha Coakley, et al. v. Bangor Hydro-Electric Co., et al., 2013, 144 FERC 63,012 at 576): "if [return on equity] is set substantially below 10% for long periods [...], it could negatively impact future investment in the (New England Transmission Owners)." Our findings here draw us back to Joskow's (1972) characterization of regulator decision-making as a sort of meta-analysis. That is, commissioners do not merely directly evaluate the CAPM equations. Rather, they look at the nature of the evidence *as presented to them*. Accordingly, their judgments are based not just on capital market conditions in a vacuum, but on the format, detail, and context of the information contained within the evidentiary record of a rate case. As a result, regulators are susceptible to biases in judgment, and calibration of regulatory decision-making during the rate-setting process should be a required step.

5. Conclusions and policy implications

In this paper, we have examined a database of electric utility rates of return authorized by U.S. state regulatory agencies over a thirty-eight-year period. These rates have demonstrated a growing spread over the riskless rate of return across the time horizon studied. The size and growth of this spread—the risk premium—does not appear to be consistent with classical finance theory, as expressed by the CAPM. In fact, regression analysis of the data suggests the *opposite* of what would be predicted if the CAPM holds. This is particularly perplexing given that regulators often *claim* to be using the CAPM. In addition to the traditional finance factors, our work examined the influence of institutional, structural, and behavioral factors on the determination of authorized rates of return. We find support for many of these factors, although most cannot be justified on traditional normative grounds.

The pattern of large and growing risk premiums illustrated in this paper has significant implications for both utility and infrastructure investment and regulation and market design in environments where both regulated and restructured firms compete for capital. In particular, if rate case activity increases over the next several years as rate moratoria expire, the implications for retail rate escalation and capital investment may be significant. We discuss each in turn before offering some thoughts on possible mitigating factors.

[17] Formally, we test the hypothesis that the observed skewness is equal to zero (a symmetric, normal distribution). The test statistic is equal to the skewness divided by its standard error $\sqrt{6n(n-1)/(n-2)(n+1)(n+3)}$, where n is the sample size. The test statistic has an approximately normal distribution (Cramer and Howitt, 2004).

[18] At the time of the 2015 survey, for example, the 10-year U.S. Treasury rate was 2.0%. The average riskless rate reported by the 1983 U.S. survey respondents was 2.4% (median 2.3%), but responses ranged from 0.0% to 8.0%.

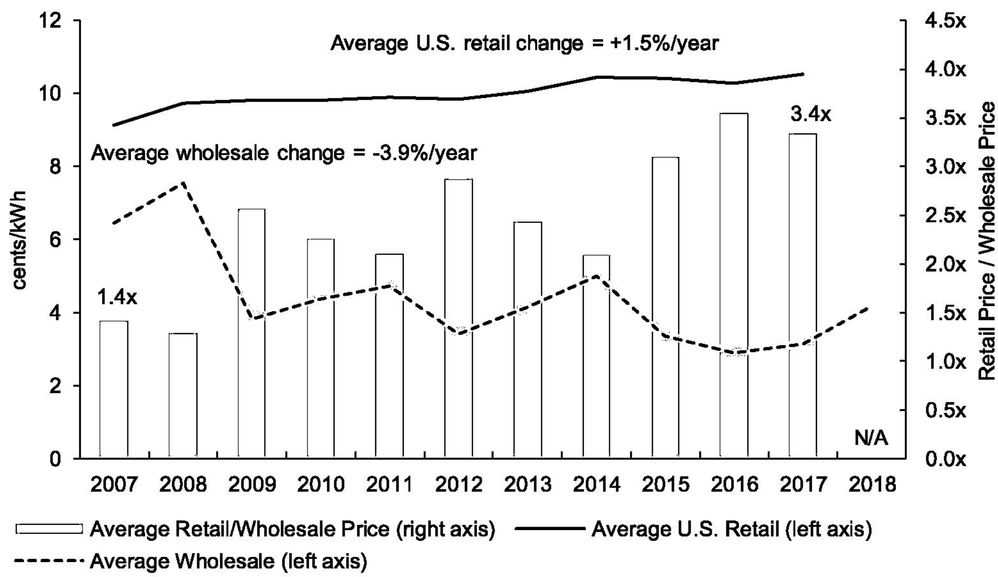


Fig. 15. Peak wholesale (2007–2018) vs. retail (2007–2017) power prices. Wholesale prices represent the average annual peak electricity price in MISO-IN, ISO-NE Mass Hub, Mid-C, Palo Verde, PJM-West, SP-15, and ERCOT-North. Retail prices collected from U.S. Energy Information Administration (https://www.eia.gov/electricity/data/state/avgprice_annual.xlsx). The retail price is the average for the entire country (using only the 7 states with wholesale markets included does not change the result).

5.1. *Wholesale and retail electricity price divergence*

A growing divergence has emerged over the last decade. Although fuel costs and wholesale power prices have declined since 2007, the *retail* price of power has increased over the same period (see Fig. 15). One explanation for this divergence in wholesale and retail rates may be the presence of a growing premium attached to regulated equity returns and therefore embedded into rates. To be sure, other forces may also be at work (for example, recovery of transmission and distribution system investments is consuming a greater portion of retail bills—a circumstance potentially exacerbated by excessive risk premiums). Further, even if the growing divergence between wholesale and retail rates is related to a growing risk premium, it does not necessarily follow that such growth is inappropriate or inconsistent with economic theory. Nevertheless, the potential for embedding of such quasi-fixed costs into the cost structure of electricity production may be significant for end users, as efficiency gains on the wholesale side are more than offset by excess costs of equity capital on the retail side.

5.2. *Regulation itself as a source of risk*

Public policy, or regulation itself, may be a causal factor in the observed behavior of the risk premium. The U.S. Supreme Court acknowledged, in *Duquesne Light Company* et al. *v. David M. Barasch* et al. (488 U.S. 299 (1989), p. 315) that "the risks a utility faces are in large part defined by the rate methodology, because utilities are virtually always public monopolies dealing in an essential service, and so relatively immune to the usual market risks." The recognition that the very act of regulating utilities subjects them to a unique class of risks may influence their cost of capital determination. And yet, if the *purpose* (or at least *a* purpose) of regulating electric utilities is to prevent these quasi-monopolists from charging excessive prices, but the *practice* of regulating them results in a higher cost of equity capital than might otherwise apply, it calls into question the role of such regulation in the first place.

Similarly, we may also question whether the hybrid regulated and non-regulated nature of the electric power sector in the U.S. plays a role as well. Has deregulation caused risk to "leak" into the regulated world because both regulated and non-regulated firms must compete for the same pool of capital? Has the presence of non-regulated market participants raised the marginal price of capital to all firms? In Section 4.4 we illustrated a shift in the trend of risk premium growth in 1999, as several U.S. markets were switching to deregulation, but further study of this question is needed.

The trajectory of public policy during the entire time period studied has been toward deregulation (beginning before 1980 with Public Utility Regulatory Policy Act, through the Natural Gas Policy Act in the 1980s, and electric industry deregulation in the 1990s) and "today's investments face market, political and regulatory risks, many of which have no historical antecedent that might serve as a starting point for modeling risk." (PJM Interconnection, 2016) The general unobservability of the *ex ante* expected returns on deregulated assets complicates determining if the progressive deregulation of the industry has caused a convergence in regulated and non-regulated returns over that time period. While the data do not suggest that utilities in states that have never undertaken deregulation have meaningfully different risk premiums, there are many ways to evaluate the "degree" of deregulatory activity that could be explored.

Another public policy-related factor could be a change in the nature of the rate base or of rate-making itself. Toward the beginning of our study period, most of the electric utilities were vertically integrated (i.e., in the business of both generation and transmission of power). Over time, generation became increasingly exposed to deregulation, while transmission and distribution of power have tended to remain regulated. To the extent that the portion of the rate base comprised of transmission and distribution assets has increased at the expense of generation assets, it may suggest a shift in the underlying risk profile of the assets being recognized by regulators. We note, for example, that public policy has tended to favor transmission investments with "incentive rates" in recent years in order to address a perceived relative lack of investment in transmission within the electric power sector. Our data, however, reveal the opposite. Based on data since 2000, there have been 172 transmission and distribution-only cases, out of 653 total cases. The average rate of return authorized in the transmission and distribution cases is approximately 60 basis points *lower* than those in vertically-integrated cases from the same period. These have been *state-*

level cases however. We note as deserving of further study that (interstate) electric transmission is regulated by FERC using a well-defined DCF approach instead of CAPM. The impact of having differing regulatory frameworks to set rates for assets that are functionally substantially identical remains an open question.

As for a change in the nature of rate-making itself, we note that the industry has tended to move from cost-of-service rate-making to performance-based ratemaking. If this shift, in an attempt to increase utility operating efficiency, has inadvertently raised the cost of equity capital through the use of incentive rates, it would be important to ascertain if the net cost-benefit balance has been positive. In general, there has been a lack of attention to the impact of regulatory changes on discount rates. The data on authorized returns on equity provides a unique dataset for such investigations.

5.3. Strategies for mitigating the growing premium

Our research does not necessarily imply that the rates of return authorized by regulators are too high, or otherwise necessarily inappropriate for utilities. An evaluation of whether these non-normative factors constitute a legitimate basis of rate of return determination deserves separate study. But if institutional or behavioral factors lead to departures from normative outcomes in setting rates of return on equity, then perhaps like Ulysses and the Sirens, regulators' hands should be "tied to the mast."

One notable jurisdictional difference in regulatory practice is between formulaic and judgment-based approaches to setting the cost of capital. In Canada, for example, formulaic approaches are more prevalent than in the United States (Villadsen and Brown, 2012). California also adjusts returns on equity for variations in bond yields beyond a "dead band," and the performance-based regulatory approaches in Mississippi and Alabama rely on formulaic cost of capital determination (Villadsen et al., 2017).

By pre-committing to a set formula (e.g., government bond rates plus *n* basis points) in lieu of holding adversarial hearings, regulators could minimize the potential for deviation from outcomes consistent with finance theory. Villadsen and Brown (2012) noted, for example, that then-recent rates set by Canadian regulators tended to be lower than those set by U.S. regulators despite nearly equivalent riskless rates of return. An intermediate approach would be to require regulators to calculate and present a formulaic result, but then allow them the discretion to authorize deviations from such a result when circumstances justify such departures. In such cases, regulators could avoid anchoring on past results, and instead anchor on a theoretically-justifiable return, before adjusting for any mitigating factors. If regulator judgment is impaired or subject to bias, then minimizing the influence of judgment by deferring to models may be prudent. In the end, we may observe simply that what regulators *should* do, what regulators *say* they're doing, and what regulators *actually* do may be three very different things.

Conflicts of interest

The authors declare that they have no conflict of interest.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgements

This paper is based on a portion of the first author's doctoral dissertation. The authors gratefully acknowledge the thoughtful comments and assistance received from Jay Apt, Tony Páez, Thomas Yu, Chad Schafer, participants at the Carnegie Mellon Electricity Industry Center's 2017 Advisory Committee Meeting, and two anonymous reviewers. All responsibility for any errors remains with the authors.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.enpol.2019.110891.

References

Advanced Energy Economy, 2018. PowerPortal. https://powersuite.aee.net/portal, Accessed date: 2 January 2019.
Andrews, D., 1993. Tests for parameter instability and structural change with unknown change point. Econometrica 61, 821–856. https://doi.org/10.2307/2951764.
Aragon, G., Dieckmann, S., 2011. Stock market trading activity and returns around milestones. J. Empir. Financ. 18, 570–584. https://doi.org/10.1016/j.jempfin.2011.05.010.
Averch, H., Johnson, L., 1962. Behavior of the firm under regulatory constraint. Am. Econ. Rev. 52, 1052–1069.
Balduzzi, P., Foresi, S., Hait, D., 1997. Price barriers and the dynamics of asset prices in equilibrium. J. Financ. Quant. Anal. 32, 137–159. https://doi.org/10.2307/2331170.
Besley, T., Coate, S., 2003. Elected versus appointed regulators: theory and evidence. J. Eur. Econ. Assoc. 1, 1176–1206. https://doi.org/10.1162/154247603770383424.
Binder, J., Norton, S., 1999. Regulation, profit variability and beta. J. Regul. Econ. 15, 249–265. https://doi.org/10.1023/A:1008078230272.
Board of Governors of the Federal Reserve System (US), n.d.-a. 10-Year Treasury Constant Maturity Rate [DGS10], Retrieved from FRED, Federal Reserve Bank of St. Louis. https://research.stlouisfed.org/fred2/series/DGS10 (accessed 2 January 2019).
Board of Governors of the Federal Reserve System (US), n.d.-b. Moody's Seasoned Baa Corporate Bond Yield [BAA], Retrieved from FRED, Federal Reserve Bank of St. Louis. https://research.stlouisfed.org/fred2/series/BAA (accessed 2 January 2019).
Borenstein, S., Bushnell, J., 2015. The U.S. electricity industry after 20 years of restructuring. Annu. Rev. of Econ. 7, 437–463. https://doi.org/10.1146/annurev-economics-080614-115630.
Boyes, W., McDowell, J., 1989. The selection of public utility commissioners: a reexamination of the importance of institutional setting. Public Choice 61, 1–13. https://doi.org/10.1007/BF00116759.
Chow, G., 1960. Tests of equality between sets of coefficients in two linear regressions. Econometrica 28, 591–605. https://doi.org/10.2307/1910133.
Cohen, R., 2008. Incorporating default risk into Hamada's equation for application to capital structure. Wilmott Mag. March/April 62–68.
Conine Jr., T., 1980. Corporate debt and corporate taxes: an extension. J. Financ. 35, 1033–1037. https://doi.org/10.1111/j.1540-6261.1980.tb03519.x.
Costello, K., 1984. Electing regulators: the case of public utility commissioners. Yale J. Regul. 2, 83–105.
Courville, L., 1974. Regulation and efficiency in the electric utility industry. Bell J. Econ. Manag. Sci. 5, 53–74. https://doi.org/10.2307/3003092.
Cramer, D., Howitt, D., 2004. The SAGE Dictionary of Statistics. SAGE Publications, Thousand Oaks, pp. 159. https://doi.org/10.4135/9780857020123.
Damodaran, A., n.d. Annual Returns on Stock, T-Bonds and T-Bills: 1928-Current. http://pages.stern.nyu.edu/~adamodar/New_Home_Page/datafile/histretSP.html (accessed 2 January 2019).
Donaldson, R., Kim, H., 1993. Price barriers in the Dow Jones industrial average. J. Financ. Quant. Anal. 28, 313–330. https://doi.org/10.2307/2331416.
Fama, E., French, K., 1993. Common risk factors in the returns on stocks and bonds. J. Financ. Econ. 33, 3–56. https://doi.org/10.1016/0304-405X(93)90023-5.
Fehr, E., Tyran, J., 2001. Does money illusion matter? Am. Econ. Rev. 91, 1239–1262. https://doi.org/10.1257/aer.91.5.1239.
Fernandez, P., Ortiz, A., Acín, I., 2015. Discount Rate (Risk-free Rate and Market Risk Premium) Used for 41 Countries in 2015: A Survey. http://ssrn.com/abstract=2598104, Accessed date: 2 January 2019.
Fernandez, P., Pershin, V., Acín, I., 2017. Discount Rate (Risk-free Rate and Market Risk Premium) Used for 41 Countries in 2017: A Survey. http://ssrn.com/abstract=2954142, Accessed date: 2 January 2019.
Fernandez, P., Pershin, V., Acín, I., 2018. Market Risk Premium and Risk-free Rate Used for 59 Countries in 2018: A Survey. http://ssrn.com/abstract=3155709, Accessed date: 2 January 2019.
Fisher, I., 1928. The Money Illusion. Adelphi, New York.
Hagerman, R., Ratchford, B., 1978. Some determinants of allowed rates of return on equity to electric utilities. Bell J. Econ. 9, 46–55. https://doi.org/10.2307/3003611.
Hamada, R., 1969. Portfolio analysis, market equilibrium and corporation finance. J. Financ. 24, 13–31. https://doi.org/10.1111/j.1540-6261.1969.tb00339.x.
Hamada, R., 1972. The effect of the firm's capital structure on the systematic risk of common stocks. J. Financ. 27, 435–452. https://doi.org/10.1111/j.1540-6261.1972.tb00971.x.
Harris, M., Navarro, P., 1983. Does electing public utility commissioners bring lower electric rates? Public Util. Fortn. 112, 23–27.
Haug, T., Wieshammer, L., 2019. Cost of equity for regulated networks: recent developments in continental Europe. Electr. J. 32, 1–3. https://doi.org/10.1016/j.tej.2019.02.001.
Hausman, J., 1978. Specification tests in econometrics. Econometrica 46, 1251–1272. https://doi.org/10.2307/1913827.
Hausman, J., Taylor, W., 1981. Panel data and unobservable individual effects.

Econometrica 49, 1377–1398. https://doi.org/10.2307/1911406.
Internal Revenue Service, n.d. SOI Tax Stats, Historical Table 24. https://www.irs.gov/pub/irs-soi/histabb.xls (accessed 2 January 2019).
Joskow, P., 1974. Inflation and environmental concern: structural change in the process of public utility price regulation. J. Law Econ. 17, 291–327. https://doi.org/10.1086/466794.
Hayashi, P., Trapani, J., 1976. Rate of return regulation and the regulated firm's choice of capital-labor ratio: further empirical evidence on the Averch-Johnson model. South. Econ. J. 42, 384–398.
Joskow, P., 1972. The determination of the allowed rate of return in a formal regulatory hearing. Bell J. Econ. Manag. Sci. 3, 632–644. https://doi.org/10.2307/3003042.
Judge, G., Griffiths, W., Hill, R., Lütkepohl, H., Lee, T., 1985. The Theory and Practice of Econometrics, second ed. John Wiley, New York.
Koedijk, K., Stork, P., 1994. Should we care? Psychological barriers in stock markets. Econ. Lett. 44, 427–432. https://doi.org/10.1016/0165-1765(94)90116-3.
Kwoka, J., 2002. Governance alternatives and pricing in the U.S. electric power industry. J. Law Econ. Organ. 18, 278–294. https://doi.org/10.1093/jleo/18.1.278.
Lintner, J., 1965. Security prices, risk, and maximal gains from diversification. J. Financ. 20, 79–96. https://doi.org/10.1111/j.1540-6261.1965.tb02930.x.
Mehra, R., Prescott, E., 1985. The equity premium: a puzzle. J. Monet. Econ. 15, 145–161. https://doi.org/10.1016/0304-3932(85)90061-3.
Mehra, R., Prescott, E., 2003. The equity premium in retrospect. In: Constantinides, G., Harris, M., Stulz, R. (Eds.), Handbook of the Economics of Finance. Elsevier, Amsterdam, pp. 889–938.
Morin, R., 2006. New Regulatory Finance. Public Utilities Reports, Vienna.
Mossin, J., 1966. Equilibrium in a capital asset market. Econometrica 34, 768–783. https://doi.org/10.2307/1910098.
Myers, S., 1972. The application of finance theory to public utility rate cases. Bell J. Econ. Manag. Sci. 3, 58–97. https://doi.org/10.2307/3003071.
Navarro, P., 1982. Public utility commission regulation: performance, determinants, and energy policy impacts. Energy J. 3, 119–139. https://doi.org/10.5547/ISSN0195-6574-EJ-Vol3-No2-7.
Olson, M., 1965. The Logic of Collective Action. Harvard University Press, Cambridge.
Pennsylvania Public Utility Commission, 2016. Bureau of Technical Utility Services Report on the Quarterly Earnings of Jurisdictional Utilities for the Year Ended June 30, 2016. Docket M-2016-2570997. . http://www.puc.state.pa.us/pcdocs/1483248.docx, Accessed date: 2 January 2019.
Pettway, R., Jordan, B., 1987. APT vs. CAPM estimates of the return-generating function parameters for regulated public utilities. J. Financ. Res. 10, 227–238. https://doi.org/10.1111/j.1475-6803.1987.tb00493.x.
Phillips Jr., C., 1993. The Regulation of Public Utilities. Public Utilities Reports, Arlington.
PJM Interconnection, 2016. Resource Investment in Competitive Markets. http://pjm.com/~/media/documents/reports/20160505-resource-investment-in-competitive-markets-paper.ashx, Accessed date: 2 January 2019.
Primeaux, W., Mann, P., 1986. Regulator selection methods and electricity prices. Land Econ. 62, 1–13. https://doi.org/10.2307/3146558.
Roberts, R., Maddala, G., Enholm, G., 1978. Determinants of the requested rate of return and the rate of return granted in a formal regulatory process. Bell J. Econ. 9, 611–621. https://doi.org/10.2307/3003601.
Roll, R., Ross, S., 1983. Regulation, the capital asset pricing model, and the arbitrage pricing model. Public Util. Fortn., May 26 (1983), 22–28.
Ross, S., 1976. The arbitrage theory of capital asset pricing. J. Econ. Theory 13, 341–360. https://doi.org/10.1016/0022-0531(76)90046-6.
Sears, R., Trennepohl, G., 1993. Investment Management. Dryden Press, Ft. Worth, pp. 516–517.
Shafir, E., Diamond, P., Tversky, A., 1997. Money illusion. Q. J. Econ. 112, 341–374. https://doi.org/10.1162/003355397555208.
Sharpe, W., 1964. Capital asset prices: a theory of market equilibrium under conditions of risk. J. Financ. 19, 425–442. https://doi.org/10.1111/j.1540-6261.1964.tb02865.x.
Spann, R., 1974. Rate of return regulation and efficiency in production: an empirical test of the Averch-Johnson thesis. Bell J. Econ. Manag. Sci. 5, 38–52. https://doi.org/10.2307/3003091.
Stigler, G., 1971. The theory of economic regulation. Bell J. Econ. Manag. Sci. 2, 3–21. https://doi.org/10.2307/3003160.
Svedsäter, H., Gamble, A., Gärling, T., 2007. Money illusion in intuitive financial judgments: influences of nominal representation of share prices. J. Socio Econ. 36, 698–712. https://doi.org/10.1016/j.socec.2007.01.006.
Villadsen, B., Brown, T., 2012. International innovations in rate of return determination. In: Presentation at the 2012 Society of Utility and Regulatory Financial Analysts Forum, April 26, 2012.
Villadsen, B., Vilbert, M., Harris, D., Kolbe, A.L., 2017. Risk and Return for Regulated Industries. Academic Press, Cambridge.
Vitaliano, D., Stella, G., 2009. A frontier approach to testing the Averch-Johnson hypothesis. Int. J. Econ. Bus. 16, 347–363. https://doi.org/10.1080/13571510903227031.

THE WALL STREET JOURNAL.

Why the Dow 36000 Forecast Was Right

Hassett and Glassman were mocked in 1998 but have been vindicated by rising stocks. The danger now? Rising interest rates.

Ken Rogoff

September 8, 2021

Journalist James Glassman and economist Kevin Hassett took to these pages in March 1998 to dismiss concerns of a stock market bubble and offer a rationale for why stocks could still go a lot higher. That was already sticking their necks out. But then they stuck them out further, arguing—with many qualifications—that over the long run the Dow Jones Industrial Average could go to 35000. Their simple policy advice: Buy a diversified portfolio of stocks, and don't put too much money in bonds. Over the long run, stocks offer a higher return without significantly greater risk. Their basic rationale, which depended on investors gradually coalescing around their view, was that stocks had to rise as the market risk premium came into line with empirical reality.

Later, in 1999, they stuck their necks out even further in a best-selling book titled "Dow 36,000." In it, their very rough guess on timing was five years. The left hated them, in no small part because Mr. Hassett is a prominent conservative economist. The right liked them, partly for the same reason. On balance, it is probably fair to say that the book was viewed as somewhere between outrageous and absurd.

Well, this summer, 22 years later, the Dow closed above 35000 for the first time, and 36000 is within reach. Is it time to admit the book might also have been a wee bit prescient? In my view, it is.

Admittedly, making that judgment is a little like deciding how to grade a student who intuited a strong answer to a very difficult exam problem but whose proof had a significant logical flaw. At the time of the Journal piece, the Dow had hit a lofty 8800, a value that had unnerved many pundits, especially considering the Dow had been 3400 only five years earlier. Indeed, when the Dow hit 6400 in December 1996, Federal Reserve chairman Alan Greenspan, worried that the financial system might be reaching a vulnerable zone, gave his legendary "irrational exuberance" speech. Mr. Greenspan was hardly alone; a few days before his speech, Yale finance expert Robert J. Shiller (a Nobel laureate in 2013) had given a presentation to the Federal Reserve Board explaining his view that there was a high risk of a stock market crash.

For years, the Glassman-Hassett analysis was widely denigrated, particularly by left-leaning commentators, who delighted in equating conservative and stupid. Nobel Prize-winning economist Paul Krugman called the idea "silly," albeit having once exposited their case for much higher stock prices quite eloquently. Berkeley economist Brad DeLong echoed a review castigating the book's "incredibly money-losing advice." And those were not the worst. I confess that I, too, strongly questioned Messrs. Glassman and Hassett's zero long-run risk logic, although I did stress that there was academic gravitas to their core point that, historically, a diversified portfolio of U.S. stocks has typically outperformed safe bonds over very long periods—say, of 25 to 30 years—even if one can construct exceptions.

Truth be told, Messrs. Glassman and Hassett's theory was "Stocks for the Long Run," Jeremy Siegel's 1994 best-seller, on steroids. Mr. Siegel in turn had extended and reinterpreted Rajnish Mehra and Edward C. Prescott's seminal 1985 paper on the "equity premium puzzle." Rather than view the consistent excess long-run real returns on stocks as a theoretical puzzle, as Messrs. Mehra and Prescott did, Mr. Siegel argued that the long-run excess return on stocks was a massive investment opportunity, offering to patient investors a much higher return on modestly higher risk. Messrs. Glassman and Hassett took the argument one step further still, basically guessing that stocks would keep rising and rising as the arguments of Messrs. Siegel, Mehra and Prescott became more widely accepted, leading investors to bid up stock prices.

So, what, if anything, did they get wrong? Well, for starters, stocks are certainly risky in the short to medium run, and for a lot of investors that is a big deal. If you have $40 million and overnight it becomes $20 million, you can probably afford to be patient, however long it takes for the market to come back. But if you are elderly with $300,000 in retirement savings, seeing it collapse to $150,000 is pretty painful, and the knowledge your investment will be worth much more if you can wait two decades won't dull the pain.

Even a long-lived foundation or university faces constraints from trustees and regulators that can effectively force an uncomfortable spending adjustment if the market collapses. The pool of players who care about short and medium returns is very substantial. The Mehra-Prescott model does not allow for liquidity constraints, which are important for a sizable chunk of the market.

Now that the Dow is approaching 36000, is it actually because people are putting more and more of their investments into risky assets? In part, yes. But surely the biggest driver of prices for all long-lived assets—including equities, housing, art and even bitcoin—has been the extraordinarily low level of real interest rates. It's not so much that the equity premium has fallen as that the interest rate on bonds has collapsed.

Nevertheless, Messrs. Glassman and Hassett also got something very right. Their extreme take on the equity premium puzzle—the mystery of why the average return on stocks seems so high relative to safe bonds—hits an important point that has operational significance for many investors. Those with the wealth and liquidity to ride out short- and medium-run fluctuations have an enormous advantage. For those with little wealth, the advice to invest in stocks is not very helpful. Yet for the majority of middle-income Americans, who certainly understand long-term investing when it comes to housing, the equity premium is something they should know about and make their own decisions.

What's next? The two-plus decades of experience since "Dow 36,000" has changed my mind far more about the equilibrium real interest rate than it has about the riskiness of stocks, including in the medium and long run. Risk markets could wilt if and when global real interest rates start trending up, or if there were a real or cyber war. There is ample reason to be nervous. Nevertheless, if Messrs. Glassman and Hassett decide to publish a new edition called "Dow 72,000," I will buy a copy.

Mr. Rogoff is a professor of economics at Harvard and a former chief economist at the International Monetary Fund.

https://www.wsj.com/articles/dow-36000-forecast-right-stocks-bonds-records-glassman-hassett-11631117174

OPINION | COMMENTARY

Why the Dow 36000 Forecast Was Right

Hassett and Glassman were mocked in 1998 but have been vindicated by rising stocks. The danger now? Rising interest rates.

By Kenneth Rogoff
Sept. 8, 2021 12:30 pm ET

ILLUSTRATION: CHAD CROWE

Journalist James Glassman and economist Kevin Hassett took to these pages in March 1998 to dismiss concerns of a stock market bubble and offer a rationale for why stocks could still go a lot higher. That was already sticking their necks out. But then they stuck them out further, arguing—with many qualifications—that over the long run the Dow Jones Industrial Average could go to 35000. Their simple policy advice: Buy a diversified portfolio of stocks, and don't put too much money in bonds. Over the long run, stocks offer a higher return without significantly greater risk. Their basic rationale, which depended on investors gradually coalescing around their view, was that stocks had to rise as the them, partly for the same reason. On balance, it is probably fair to say that the book was viewed as somewhere between outrageous and absurd.

Well, this summer, 22 years later, the Dow closed above 35000 for the first time, and 36000 is within reach. Is it time to admit the book might also have been a wee bit prescient? In my view, it is.

Admittedly, making that judgment is a little like deciding how to grade a student who intuited a strong answer to a very difficult exam problem but whose proof had a significant logical flaw. At the time of the Journal piece, the Dow had hit a lofty 8800, a value that had unnerved many pundits, especially considering the Dow had been 3400 only five years earlier. Indeed, when the Dow hit 6400 in December 1996, Federal Reserve chairman Alan Greenspan, worried that the financial system might be reaching a vulnerable zone, gave his legendary "irrational exuberance" speech. Mr. Greenspan was hardly alone; a few days before his speech, Yale finance expert Robert J. Shiller (a Nobel laureate in 2013) had given a presentation to the Federal Reserve Board explaining his view that there was a high risk of a stock market crash.

For years, the Glassman-Hassett analysis was widely denigrated, particularly by left-leaning commentators, who delighted in equating conservative and stupid. Nobel Prize-winning economist Paul Krugman called the idea "silly," albeit having once exposited their case for much higher stock prices quite eloquently. Berkeley economist Brad DeLong outperformed safe bonds over very long periods—say, of 25 to 30 years—even if one can construct exceptions.

Truth be told, Messrs. Glassman and Hassett's theory was "Stocks for the Long Run," Jeremy Siegel's 1994 best-seller, on steroids. Mr. Siegel in turn had extended and reinterpreted Rajnish Mehra and Edward C. Prescott's seminal 1985 paper on the "equity premium puzzle." Rather than view the consistent excess long-run real returns on stocks as a theoretical puzzle, as Messrs. Mehra and Prescott did, Mr. Siegel argued that the long-run excess return on stocks was a massive investment opportunity, offering to patient investors a much higher return on modestly higher risk.

Messrs. Glassman and Hassett took the argument one step further still, basically guessing that stocks would keep rising and rising as the arguments of Messrs. Siegel, Mehra and Prescott became more widely accepted, leading investors to bid up stock prices.

So, what, if anything, did they get wrong? Well, for starters, stocks are certainly risky in the short to medium run, and for a lot of investors that is a big deal. If you have $40 million and overnight it becomes $20 million, you can probably afford to be patient, however long it takes for the market to come back. But if you are elderly with $300,000 in retirement savings, seeing it collapse to $150,000 is pretty painful, and the knowledge your investment will be worth much more if you can wait two decades won't dull the pain.

Even a long-lived foundation or university faces constraints from trustees and regulators that can effectively force an uncomfortable spending adjustment if the market collapses. The pool of players who care about short and medium returns is very substantial. The Mehra-Prescott model does not allow for liquidity constraints, which are important for a sizable chunk of the market.

Now that the Dow is approaching 36000, is it actually because people are putting more and more of their investments into risky assets? In part, yes. But surely the biggest driver of prices for all long-lived assets—including equities, housing, art and even bitcoin—has been the extraordinarily low level of real interest rates. It's not so much that the equity premium has fallen as that the interest rate on bonds has collapsed.

medium-run fluctuations have an enormous advantage. For those with little wealth, the advice to invest in stocks is not very helpful. Yet for the majority of middle-income Americans, who certainly understand long-term investing when it comes to housing, the equity premium is something they should know about and make their own decisions.

What's next? The two-plus decades of experience since "Dow 36,000" has changed my mind far more about the equilibrium real interest rate than it has about the riskiness of stocks, including in the medium and long run. Risk markets could wilt if and when global real interest rates start trending up, or if there were a real or cyber war. There is ample reason to be nervous. Nevertheless, if Messrs. Glassman and Hassett decide to publish a new edition called "Dow 72,000," I will buy a copy.

Mr. Rogoff is a professor of economics at Harvard and a former chief economist at the International Monetary Fund.

Appeared in the September 9, 2021, print edition.

Journal of Financial Economics 12 (1983) 371-386. North-Holland

ON COMPUTING MEAN RETURNS AND THE SMALL FIRM PREMIUM

Richard ROLL*

University of California, Los Angeles, CA 90024, USA

Received January 1983, final version received June 1983

The mean return computational method has a substantial effect on the estimated small firm premium. The buy-and-hold method, which best mimics actual investment experience, produces an estimated small-firm premium only one-half as large as the arithmetic and re-balanced methods which are often used in empirical studies. Similar biases can be expected in mean returns when securities are classified by any variable related to trading volume.

1. Introduction

There is a potentially serious problem in estimating expected return differences between small and large firms. Even with exactly the same sample observations, the method used to compute sample mean returns can have a substantial effect on the estimates.

With an arithmetic computational method, daily returns on individual stocks are averaged across both firms and days to obtain the mean daily return on an equally-weighted portfolio; then the portfolio's mean daily return is compounded to obtain an estimate of the expected return over a longer interval. With a buy-and-hold method, individual stock returns are first obtained for the longer interval by linking together the daily individual returns; then an equally-weighted portfolio's mean return is computed by averaging the longer-term (individual) returns.

Defining a 'longer interval' as one year, the arithmetic method produces an average annual return difference of 14.9 percent between AMEX and NYSE stocks[1] over the 19 complete calendar years, 1963–1981 inclusive. The buy-and-hold method gives an annual return difference of only 7.45 percent. Assuming that annual returns are statistically independent, the arithmetic

*Comments and suggestions by Gordon Alexander, Kenneth French, Stephen Ross and the referee, Allan Kleidon, are gratefully acknowledged.

[1]The effect of smallness can be measured by the difference in returns of stock listed on the American Exchange (AMEX) and the New York Exchange (NYSE) because AMEX issues are, on average, much smaller than NYSE issues. Most of the results presented here are based on the AMEX–NYSE differential because it is convenient and easy to use. Some confirmatory results based directly on measured size will also be presented.

method's return differential had an associated t-statistic of 3.07 while the buy-and-hold method yielded a t-statistic of 1.53.

Speculation on possible causes of the small firm premium has occupied the attention of many finance theorists over the past few years; but perhaps this attention has been premature. If the estimated small firm premium can be cut in half simply by compounding individual returns before averaging them, some consideration should be given to whether the magnitude of the true premium is really all that large. The various explanations for the premium offered so far would become more plausible if the premium is actually smaller than has been previously reported.

This paper investigates why the mean return computational method can be such a significant choice in some empirical research. The reason seems to be that individual asset returns are not as well-behaved as we might like. Individual assets do not trade continuously and there are significant trading costs. In some empirical studies, the effect of these factors might be safely ignored; but when the object of investigation is related to trading volume (and thus to trading frequency and trading costs), there can be measurement problems. Firm size is related to trading volume and it is used as an example throughout the paper. Other variables related to size and to trading, such as dividend yield, price/earnings ratio, and beta, could also present similar empirical difficulties. Section 2 gives a brief theoretical discussion of mean return computational methods and section 3 presents details of the empirical results for small firm premia.

2. Compounding and the bias in mean return calculation

2.1. Formulae for computing mean returns

To elucidate the differences in mean return computation and explain why they might produce different results, consider a sample of N securities, each having returns observed for T periods. Let R_{it} be the value relative (1+return), of security i in period t. Suppose also that investment results are reviewed every τ periods. For example, if data were available daily but returns were to be reviewed every month, we would have $\tau \cong 21$ since there are usually about 21 trading days per month.

Two alternative methods of computing the mean equally-weighted return over the review period can be written algebraically as

$$\bar{R}_{\mathrm{AR}} = \left[\frac{1}{N \cdot \tau} \sum_i \sum_t R_{it}\right]^{\tau}, \tag{1}$$

$$\bar{R}_{\mathrm{BH}} = \frac{1}{N} \sum_i \left[\prod_t R_{it}\right], \tag{2}$$

where the subscripts 'AR' and 'BH' denote 'arithmetic' and 'buy-and-hold', respectively. These labels are intended to portray the sense of the computation method. The first method (1) is simply an arithmetic mean raised to the τth power while the second method gives the actual investment results an investor would achieve from buying equal dollar amounts of N securities and holding the shares for τ periods.

There is also a third possible definition of mean return,

$$\bar{R}_{\mathrm{RB}} = \prod_t \left[\frac{1}{N} \sum_i R_{it} \right], \tag{3}$$

where the subscript 'RB' stands for 'rebalanced'. This would be the actual investment return (ignoring transactions costs) on a portfolio which begins with equal investments in the N securities and *maintains* equal investments by rebalancing at the end of each period, $t = 1, \ldots, \tau$.

To compare results over different review periods, we must choose some typical and familiar calendar interval, say a year, and express the results as percentage returns over that common calendar interval. In the tables below, annualization is accomplished and reported for 'linked' returns; the review period returns within each calendar year are simply multiplied together (or linked) in order to obtain an annual return.[2] Linked annualization includes *every* daily observation in some review period during the year. This assures that in any comparison of the results across review periods, the observed differences are due to review period alone and cannot be ascribed to slightly different sample observations.

The next two subsections investigate some properties of these sample mean returns. Subsection 2.2 derives their expected values under the assumption of temporally independent individual asset returns. Subsection 2.3 then examines the effect of intertemporal dependence.

[2]The exact formulae for linked returns can be written as follows. Let $\bar{R}_m(y, \tau)$ denote the mean annualized linked return for year y ($y = 1, \ldots, Y$) using a review period whose length is τ trading days and using method ($m = \mathrm{BH, AR, RB}$), to compute the review period returns. Then,

$$\bar{R}_{\mathrm{BH}}(y,\tau) = \prod_{j=(y-1)k_\tau+1}^{y \cdot k_\tau} \left[\frac{1}{N} \sum_i \prod_{t=(j-1)\tau+1}^{j \cdot \tau} (R_{it}) \right],$$

$$\bar{R}_{\mathrm{AR}}(y,\tau) = \prod_{j=(y-1)k_\tau+1}^{y \cdot k_\tau} \left[\frac{1}{N \cdot \tau} \sum_i \sum_{t=(j-1)\tau+1}^{j \cdot \tau} R_{it} \right]^\tau,$$

$$\bar{R}_{\mathrm{RB}}(y,\tau) = \prod_{j=(y-1)k_\tau+1}^{y \cdot k_\tau} \left\{ \prod_{t=(j-1)\tau+1}^{j \cdot \tau} \left[\frac{1}{N} \sum_i R_{it} \right] \right\},$$

where $k_\tau = T/(Y \cdot \tau)$ is the number of review periods per year and T is the total number of trading days in the entire sample. When returns are reviewed in natural calendar intervals such as months, the review period cannot be a fixed number of trading days and thus τ in the formulae above varies slightly with the actual number of trading days.

2.2. Sample mean return biases with temporal independence

Following Blume (1974), assume that each individual asset return is drawn from a stationary distribution with temporally independent disturbances; that is,

$$\tilde{R}_{it} = \mu_i + \tilde{\varepsilon}_{it}, \quad \forall i, \tag{4}$$

with $\mathrm{E}(\tilde{R}_{it}) = \mu_i$, a constant for all t, and where the unexpected return, $\tilde{\varepsilon}_{it}$, satisfies $\mathrm{cov}(\tilde{\varepsilon}_{i,t}, \tilde{\varepsilon}_{i,t-j}) = 0$ for $j \neq 0$.

The expected value of the arithmetic mean (1) can be expressed as

$$\mathrm{E}(\bar{R}_{\mathrm{AR}}) = \mathrm{E}\left[\left(\frac{1}{N}\sum_i \mu_i + \tilde{h}\right)^{\tau}\right], \tag{5}$$

where

$$\tilde{h} \equiv \frac{1}{N \cdot \tau} \sum_i \sum_t \tilde{\varepsilon}_{it}$$

is the average disturbance on the equally-weighted portfolio over the sample review period τ.

The expected value of the buy-and-hold mean (2) is

$$\mathrm{E}(\bar{R}_{\mathrm{BH}}) = \frac{1}{N}\sum_i \left[\mathrm{E}\prod_t (\mu_i + \tilde{\varepsilon}_{it})\right] = \frac{1}{N}\sum_i (\mu_i^{\tau}). \tag{6}$$

This follows since the expectation can be taken inside the product with independent returns and since $\mathrm{E}(\tilde{\varepsilon}) = 0$, by definition.

The rebalancing method (3) produces a mean return whose expectation is

$$\mathrm{E}(\bar{R}_{\mathrm{RB}}) = \prod_t \left[\frac{1}{N}\sum_i \mu_i\right] = \left(\frac{1}{N}\sum_i \mu_i\right)^{\tau}, \tag{7}$$

where, again, the expectation can be taken inside the product because of time independence.

Expressions (5), (6) and (7) imply that the three different mean return definitions do not produce the same results. By Jensen's inequality,

$$\mathrm{E}(\bar{R}_{\mathrm{AR}}) \geqq \mathrm{E}(\bar{R}_{\mathrm{RB}}),^{3}$$

[3] Jensen's inequality for a random variable $\tilde{x}$ and a convex function $f(x)$ is $\mathrm{E}[f(\tilde{x})] \geqq f[\mathrm{E}(x)]$. Let $\tilde{x} \equiv (1/N)\sum_i \mu_i + \tilde{h}$; then $f(\tilde{x}) = \tilde{x}^{\tau}$ is convex since $\tau > 1$. $\mathrm{E}(\bar{R}_{\mathrm{AR}}) > \mathrm{E}(\bar{R}_{\mathrm{RB}})$ follows immediately from (5) and (7) since $\mathrm{E}(\tilde{h}) = 0$.

with strict inequality if $\operatorname{var}(\tilde{h})>0$, and

$$\mathrm{E}(\bar{R}_{\mathrm{BH}}) \geqq \mathrm{E}(\bar{R}_{\mathrm{RB}}),^{4}$$

with strict inequality if $N>1$ and at least two assets have different returns. Since we generally have some randomness $[\operatorname{var}(\tilde{h})>0]$, and many securities, $(N>1)$, the rebalanced method generally should produce lower mean returns than either the arithmetic or the buy-and-hold method, provided that returns are temporally independent.

The relation between the buy-and-hold and arithmetic means is more complex; and, indeed, neither is invariably smaller than the other. The larger the cross-sectional dispersion of individual expected returns, the larger $\mathrm{E}(\bar{R}_{\mathrm{BH}})$ relative to $\mathrm{E}(\bar{R}_{\mathrm{AR}})$. But there is an offsetting influence: the larger the intertemporal dispersion of unexpected returns $(\tilde{h})$, the larger $\mathrm{E}(\bar{R}_{\mathrm{AR}})$ relative to $\mathrm{E}(\bar{R}_{\mathrm{BH}})$.[5] Their relation in a given sample depends, therefore, on the characteristics of the underlying individual returns.

2.3. Time series dependence and its effect on estimated expected returns

The effect of serial dependence is seen most easily by examining expected mean returns when the review period is doubled, say from daily to bi-daily or from bi-weekly to monthly. Assume first that returns are collected for the shorter review period and then let $\tau=2$ (a doubling of the period). Over the doubled review period, the three mean returns are

$$\bar{R}_{\mathrm{AR}}=\left[\frac{1}{N}\sum_{i}\left(\mu_i+\frac{\varepsilon_{i1}+\varepsilon_{i2}}{2}\right)\right]^2, \tag{8}$$

[4]Define $f(\mu_i)=\mu_i^{\tau}$, a convex function for $\tau>1$. With $1/N$ used as a (pseudo) probability, $\mathrm{E}(\bar{R}_{\mathrm{BH}})\geqq\mathrm{E}(\bar{R}_{\mathrm{RB}})$ follows immediately from (6) and (7). (Cf. footnote 3.) Strict inequality holds if at least two μ_i's are different. [This result was noted by Cheng and Deets in (1971).]

The inequality above grows with the cross-sectional dispersion in μ_i, ceteris paribus. To prove this, expand μ_i^{τ} in a Taylor series about $\bar{\mu}\equiv(1/N)\sum_i\mu_i$; the second-order term is a positive function of the cross-sectional variance in μ_i. If μ_i were cross-sectionally normally distributed, the variance *alone* would determine the size of the inequality.

[5]This can be confirmed by using a Taylor series expansion of $\mathrm{E}(\bar{R}_{\mathrm{AR}})$. Define $\bar{\mu}=(1/N)\sum_i\mu_i$; then

$$\mathrm{E}(\bar{R}_{\mathrm{AR}})=\bar{\mu}^{\tau}\,\mathrm{E}\left[1+\frac{\tilde{h}^2}{2}(\tau)(\tau-1)\bar{\mu}^{-2}+\frac{\tilde{h}^3}{3!}(\tau)(\tau-1)(\tau-2)\bar{\mu}^{-3}+\ldots+\tilde{h}^{\tau}\bar{\mu}^{-\tau}\right].$$

Jensen's inequality (see footnote 4 above), implies that $\mathrm{E}(\bar{R}_{\mathrm{BH}})>\bar{\mu}^{\tau}$ with the inequality being larger the larger the cross-sectional variance in μ_i. But the term in brackets just above shows that $\mathrm{E}(\bar{R}_{\mathrm{AR}})$ increases with the higher moments of $\tilde{h}$ (since $\bar{\mu}$ is strictly positive). For example, the second term in brackets involves the variance of $\tilde{h}$. Conceivably, this term could more than offset the cross-sectional variance in μ_i. If the unexpected arithmetic portfolio return $\tilde{h}$ happens to be normally-distributed, the expression above simplifies to $\mathrm{E}(\bar{R}_{\mathrm{AR}})=\bar{\mu}^{\tau}[1+k\cdot\operatorname{var}(\tilde{h})]$ with the constant $k>0$. In this case, there is a simple and direct tradeoff between the cross-sectional variance in expected return, μ_i, and the variance of the unexpected portfolio return, $\tilde{h}$.

$$\bar{R}_{\mathrm{BH}}=\frac{1}{N}\sum_{i}[(\mu_i+\varepsilon_{i1})(\mu_i+\varepsilon_{i2})], \tag{9}$$

$$\bar{R}_{\mathrm{RB}}=\left[\frac{1}{N}\sum_{i}(\mu_i+\varepsilon_{i1})\right]\left[\frac{1}{N}\sum_{i}(\mu_i+\varepsilon_{i2})\right], \tag{10}$$

where $R_{it}=\mu_i+\varepsilon_{it}$ is the observed return on individual stock i $(i=1,\dots,N)$ in period t, and μ_i is i's single-period (i.e., shorter review period) expected return.

For notational convenience, define the cross-sectional averages

$$\bar{\mu}=\frac{1}{N}\sum_{i}\mu_i \quad \text{and} \quad \bar{\varepsilon}_t=\frac{1}{N}\sum_{i}\varepsilon_{it}.$$

Then the three mean returns have expected values,

$$\mathrm{E}(\bar{R}_{\mathrm{AR}})=\bar{\mu}^2+\tfrac{1}{2}(\sigma_{\bar{\varepsilon}}^2+\sigma_{\bar{\varepsilon}_1,\bar{\varepsilon}_2}), \tag{11}$$

$$\mathrm{E}(\bar{R}_{\mathrm{BH}})=\frac{1}{N}\sum_{i}\mu_i^2+\frac{1}{N}\sum_{i}\sigma_{\varepsilon_{i1},\varepsilon_{i2}}, \tag{12}$$

$$\mathrm{E}(\bar{R}_{\mathrm{RB}})=\bar{\mu}^2+\sigma_{\bar{\varepsilon}_1,\bar{\varepsilon}_2}, \tag{13}$$

where σ_x^2 is the variance of x and $\sigma_{x,y}$ is the covariance of x and y.

Even with serial dependence, the expected arithmetic mean still exceeds the expected rebalanced mean in all circumstances since,

$$\mathrm{E}(\bar{R}_{\mathrm{AR}}-\bar{R}_{\mathrm{RB}})=\tfrac{1}{2}(\sigma_{\bar{\varepsilon}}^2-\sigma_{\bar{\varepsilon}_1,\bar{\varepsilon}_2})>0. \tag{14}$$

Comparing the buy-and-hold means and the rebalanced means, we have

$$\mathrm{E}(\bar{R}_{\mathrm{BH}}-\bar{R}_{\mathrm{RB}})=\sigma_{\mu_i}^2+\left(\frac{1}{N}\sum_{i}\sigma_{\varepsilon_{i1},\varepsilon_{i2}}-\sigma_{\bar{\varepsilon}_1,\bar{\varepsilon}_2}\right).$$

With no serial dependence in the ε's, the term in parentheses is zero and the BH mean would exceed the RB mean by the cross-sectional variance in expected individual returns.

However, with negative serial dependence in unexpected individual returns (ε_{i1} and ε_{i2}) or positive dependence in portfolio returns ($\bar{\varepsilon}_1$ and $\bar{\varepsilon}_2$), the rebalanced mean would become larger; enough such dependence could conceivably render it larger that the buy-and-hold mean. Since the expected arithmetic mean exceeds the expected rebalanced mean, it too could be larger than the BH mean with enough serial dependence of the right type.

There is some reason to anticipate just this type of serial dependence because of the intertemporal characteristics of individual returns. Scholes and Williams (1977, pp. 313–314) explain that because of non-synchronous trading individual assets display first-order *negative* serial dependence while diversified portfolios display *positive* dependence. A difference in the sign of serial dependence between individual assets and portfolios is relevant here because buy-and-hold (BH) means are mainly affected by individual asset serial dependence [see (12)], while the arithmetic (AR) and rebalanced (RB) means are affected by portfolio serial dependence [see (11) and (13)]. The Scholes/Williams explanation implies that BH means would tend to fall as review period lengthens while the AR and RB means would tend to rise.

There is also negative serial dependence induced in very short-term returns because of the institutional arrangement of trading. Neiderhoffer and Osborne (1966) pointed out that negative serial dependence should be anticipated when a market maker is involved in most transactions (because successive transactions are conducted at either the bid or the asked price).[6]

First-order negative serial dependence in individual returns has the effect of widening the disparity between the buy-and-hold mean and the arithmetic and rebalanced means as the review period lengthens. This follows from the fact that a doubling of the review period introduces serial covariance terms in *addition* to those already present. However, the marginal effect of lengthening the review period should probably diminish as the review period becomes longer; the effect on measured mean return should be greater when changing from, say, a daily to a weekly review period than from a monthly to an annual period. The exact impact of serial dependence can, of course, only be determined empirically and we now turn to an examination of the data.

3. The empirical small firm premium

3.1. Results

In the previous section, we found that the computational formula for sample mean returns can affect the estimated expected return. The buy-and-hold (BH) mean (2) gives an unbiased estimate of the holding period return on a realistic portfolio. The rebalanced (RB) mean (3), gives an unbiased estimate of return for its strategy but it is not realistic if the period is short since rebalancing is so costly. Except under a fortuitous combination of circumstances, the arithmetic (AR) mean (3) gives a biased estimate of *both* the rebalanced and the buy-and-hold investment returns.

[6]A paper by Blume and Stambaugh (1983), which came to my attention after the first version of this paper was written, investigates this explanation for serial dependence in detail. They find empirical results very similar to those reported here. See also Cohen et al. (1979).