A Catering Theory of Dividends*

Malcolm Baker
Harvard Business School and NBER
mbaker@hbs.edu

Jeffrey Wurgler
NYU Stern School of Business
jwurgler@stern.nyu.edu

November 7, 2002

Abstract

We develop a theory in which the decision to pay dividends is driven by investor demand. Managers cater to investors by paying dividends when investors put a stock price premium on payers and not paying when investors prefer nonpayers. To test this prediction, we construct four time series measures of the investor demand for dividend payers. By each measure, nonpayers initiate dividends when demand for payers is high. By some measures, payers omit dividends when demand is low. Further analysis confirms that the results are better explained by the catering theory than other theories of dividends.

* We would like to thank Viral Acharya, Raj Aggarwal, Katharine Baker, Randy Cohen, Gene D'Avolio, Steve Figlewski, Xavier Gabaix, Paul Gompers, Florian Heider, Dirk Jenter, Kose John, Steve Kaplan, John Long, Asis Martinez-Jerez, Colin Mayer, Holger Mueller, Eli Ofek, Lubos Pastor, Lasse Pedersen, Gordon Phillips, Raghu Rau, Jay Ritter, Rick Ruback, David Scharfstein, Hersh Shefrin, Andrei Shleifer, Erik Stafford, Jeremy Stein, Ryan Taliaferro, Jerold Warner, Luigi Zingales and seminar participants at Dartmouth, Harvard Business School, London Business School, LSE, MIT, NYU, Oxford, the University of Chicago, the University of Michigan, the University of Rochester, and Washington University for helpful comments; John Long and Simon Wheatley for data; and Ryan Taliaferro for superb research assistance. Baker gratefully acknowledges financial support from the Division of Research of the Harvard Business School.

A Catering Theory of Dividends

Abstract

We develop a theory in which the decision to pay dividends is driven by investor demand. Managers cater to investors by paying dividends when investors put a stock price premium on payers and not paying when investors prefer nonpayers. To test this prediction, we construct four time series measures of the investor demand for dividend payers. By each measure, nonpayers initiate dividends when demand for payers is high. By some measures, payers omit dividends when demand is low. Further analysis confirms that the results are better explained by the catering theory than other theories of dividends.

I. Introduction

Miller and Modigiliani (1961) prove that dividend policy is irrelevant to stock price in perfect and efficient capital markets. In that setup, no rational investor has a preference between dividends and capital gains. Arbitrage ensures that dividend policy is irrelevant.

Over forty years later, the only assumption in this proof that has not been thoroughly scrutinized is market efficiency.[1] In this paper, we present a theory of dividends that relaxes this assumption. It has three basic ingredients. First, for either psychological or institutional reasons, some investors have an uninformed, time-varying demand for dividend-paying stocks. Second, arbitrage fails to prevent this demand from driving apart the prices of stocks that do and do not pay dividends. Third, managers *cater to investor demand* – paying dividends when investors put a higher price on the shares of payers, and not paying when investors prefer nonpayers. We formalize this catering theory of dividends in a simple model.

The catering theory differs from the standard view of the effect of investor demand on dividend policy. The standard view emphasizes the irrelevance of dividend policy to share prices even when some investor clienteles have a rational preference for dividends. For example, Black and Scholes (1974) write: "If a corporation could increase its share price by increasing (or decreasing) its payout ratio, then many corporations would do so, which would saturate the demand for higher (or lower) dividend yields, and would bring about an equilibrium in which marginal changes in a corporation's dividend policy would have no effect on the price of its stock" (p. 2). This equilibrium intuition for dividend irrelevance can also be found in corporate finance textbooks.

[1] Allen and Michaely (2002) provide a comprehensive survey of payout policy research.

The catering theory and the clientele equilibrium theory differ on several key points. One is that catering takes seriously the possibility that investor demand for dividends is affected by sentiment. This adds a new and unexplored source of demand to the rational dividend clienteles considered by Black and Scholes. Another difference is that the catering view focuses more on the demand for *shares* that pay dividends, whereas the determinate supply response in a clientele equilibrium view is the *overall level* of dividends. For example, we discuss the possibility that managers cater to investors who categorize dividend-paying shares more or less together, and pay less attention to whether the yield on a particular share is three or four percent.

But perhaps the most crucial difference is that catering takes a less extreme view on how fast managers or arbitrageurs eliminate an emerging dividend premium or discount. According to Black and Scholes, managers compete so aggressively that a nontrivial dividend premium or discount never arises, and so for a given firm dividend policy remains effectively irrelevant. This argument is compelling only if fluctuations in the demand for dividends are small relative to the capacity of firms to adjust supply. It is not obvious a priori that this is the case, particularly if demand is affected by sentiment. The catering theory acknowledges the possibility of a nontrivial dividend premium, and thus the relevance of dividend policy.

The main prediction of the catering theory is that the propensity to pay dividends depends on a measurable dividend premium in stock prices. To test this hypothesis, we construct four time series measures of the demand for dividend-paying shares. The broadest one is what we simply call the dividend premium – it is the difference between the average market-to-book ratio of dividend payers and nonpayers. The other measures are the difference in the prices of Citizens Utilities' cash dividend and stock dividend share classes (between 1956 and 1989 CU had two classes of shares which differed in the form but not the level of their payouts); the average announcement effect of recent dividend initiations; and the difference between the *future* stock returns of payers and nonpayers. Intuition suggests that the dividend premium, the CU dividend premium, and initiation effects would be positively related to investor demand for dividends. In contrast, the difference in future returns of payers and nonpayers would be negatively related to any such demand – if demand for payers is so high that they are relatively overpriced, their future returns will be relatively low.

We then use these four measures of demand to explain time variation in aggregate rate of dividend initiation and omission. The results on initiations are the strongest. Each of the four demand measures is a significant predictor of the rate of initiation. The lagged dividend premium variable by itself explains a remarkable sixty percent of the annual variation in the initiation rate. Another perspective is future stock returns. When the initiation rate increases by one standard deviation, returns on payers are lower than nonpayers by nine percentage points per year over the next three years. Conversely, the omission rate increases when the dividend premium is low, and when future returns on payers are high.

After considering several alternative explanations, we conclude that the results are best explained by catering. Explanations based on time-varying firm characteristics such as investment opportunities or profitability, for example, do not account for the results: The dividend premium variable helps to explain the residual "propensity to initiate" dividends that remains after controlling for changing firm characteristics, including investment opportunities, profits, and firm size using the methodology of Fama and French (2001). Alternative explanations based on time-varying contracting problems, such as agency or asymmetric information theories, do not address many aspects of the results, for instance why dividend policy would related to the CU dividend premium and future returns. The lack of a compelling alternative explanation, plus the close connection between the predictions of catering and the patterns that we document, favors the catering explanation.

We then investigate which source of investor demand creates the time-varying dividend premium that attracts caterers. One possibility is rational dividend clienteles based on taxes, transaction costs, or institutional investment constraints. We would expect such clienteles to be satisfied by changes in the overall level of dividends, rather than the number of shares that pay dividends. The evidence does not support this prediction – initiations and omissions are related to the dividend premium, but the aggregate dividend yield, the aggregate payout ratio, or the aggregate rate of dividend *increases* are not. Moreover, the relationship between initiations and omissions and the dividend premium is also apparent after controlling for plausible proxies for rational clienteles. Another possibility is that demand is driven by investor sentiment. Consistent with this hypothesis, we find a strong correlation between the dividend premium and the closed-end fund discount.

In summary, we develop and test a catering theory of dividends that relaxes the market efficiency assumption of the M&M dividend irrelevance proof. The theory rounds out the collection of theories that relax other assumptions of the proof, and adds to the literature of behavioral corporate finance. In an early contribution, Shefrin and Statman (1984) develop theories of investor preference for dividends based on self-control problems, prospect theory, and regret aversion. The current paper is closer to recent research that views managerial decisions as rational responses to security mispricing. For example, Baker and Wurgler (2000) and Baker, Greenwood, and Wurgler (2002) view security issuance decisions as responses to mispricing or perceived mispricing, and Baker and Wurgler (2002a) develop this into a market timing theory of capital structure that relaxes the market efficiency assumption of the M&M capital structure irrelevance proof. Shleifer and Vishny (2002) develop a theory of mergers based on rational responses to mispricing. Morck, Shleifer, and Vishny (1990), Stein (1996), Baker, Stein, and Wurgler (2001), and Polk and Sapienza (2001) study rational corporate investment in inefficient capital markets. Graham and Harvey (2001) and Jenter (2001) provide more evidence that managers react to mispricing.

Section II develops the catering theory. Section III presents the main empirical results. Section IV considers alternative explanations. Section V discusses the source of investor demand for dividends. Section VI concludes.

II. A catering theory of dividends

The theory has three ingredients. First, there is a time-varying, uninformed demand for the shares of firms that pay cash dividends. Second, limits on arbitrage allow this demand to affect prices. Third, managers rationally cater in response. After discussing these ingredients, we combine them in a simple model.

A. Investor demand for dividends

We posit that at some times investors generally prefer stocks that pay cash dividends, and other time generally prefer nonpayers. A useful framework for developing this hypothesis is categorization. Categorization refers to the pervasive cognitive process of grouping objects into discrete categories such as “birds” or “chairs.” This allows related objects to be considered together, in terms of a small set of common features that define category membership, rather than as individual objects, each with its own list of identifying attributes. Categorization thus speeds up communication and inference. Rosch (1978) provides a detailed review.

In standard theory, investors do not categorize. Instead, they identify each security with a list of abstract statistics, such as mean return, variance, and covariance. In reality, as Barberis and Shleifer (2002) point out, investors often do categorize securities into "small stocks," "value stocks," "tech stocks," "old-economy stocks," "junk bonds," "utilities," and so forth. For many investors, these labels appear to capture all they want to know, or have the ability to process, about the securities within the category.

There are several reasons to suspect that certain investors and institutions categorize "dividend payers" directly or use dividends to classify stocks as "old economy," for example. Whether a stock pays dividends is clearly a salient characteristic, perhaps even more so than industry, size, or index membership, and the financial press often categorizes firms according to dividend payment.[2] The fact that many firms pay small but nonzero dividends suggests that there is a discrete component to attracting attention through dividends.

One reason why dividends are salient is a belief that dividend-paying stocks are less risky.[3] This notion is common in the popular financial press, and was once common in the academic literature – Graham and Dodd (1951) and Gordon (1959) are recognized for this idea, but Miller and Modigliani (1961) cite a number of other papers of this vintage that make the same argument. Naïve investors, such as retirees and those who hold dividend-paying stocks for "income" despite the tax penalty, would seem especially likely to fall prey to this bird-in-the-hand argument. For them, the quarterly dividend check is much more salient than daily gyrations in the stock price. If the risk tolerance of bird-in-the-hand investors changes over time, their

[2] For example, a July 16, 2002 *Wall Street Journal* article titled "Where should you invest now?" categorizes appealing investment options into TIPs, Ginnie Maes, Real Estate, and Dividend Paying Stocks. Quoting from the article: "As of Friday, prices for dividend-paying stocks in the S&P 500 stock index had fallen 8.04% vs. a loss of 28.18% for stocks in the index without dividends."

[3] Hyman (1988) describes investor reaction to Consolidated Edison's 1974 dividend omission. "It smashed the keystone of faith for investment in utilities: that the dividend is safe and will be paid." (p. 109).

preferences for payers and nonpayers will also change. This is one possible mechanism by which unsophisticated investors may display a time-varying sentiment for payers.

Another way dividends can become salient is if investors use them to infer managers' investment plans. For example, investors may interpret nonpayment, controlling for profitability, as evidence that the firm thinks it has excellent investment opportunities. Conversely, dividends may be taken as evidence that opportunities are weak. These inferences create another channel though which payers and nonpayers become categories, and suggest a second realistic mechanism to generate a time-varying sentiment across categories. That is, when investors' perceptions of overall growth opportunities are high, they prefer nonpayers, and vice-versa. Note that time variation is driven here by perceptions of growth opportunities, not risk tolerance as above. One popular model (Shiller (1984, 2000)) that combines both effects is that steady dividends mean "old-economy." Old-economy stocks are viewed as safer but also as having less potential than the "new-economy" stocks which plow back everything to finance growth.

Black and Scholes (1974) and Allen, Bernardo, and Welch (2000), among others, suggest that institutional frictions also lead to the rational categorization of payers by dividend clienteles. Imperfections that have been proposed to cause clienteles include transaction costs, taxes, and institutional investment constraints. Many endowed institutions are restricted to spending from income, for example, a clear reason to categorize payers. Others may take dividends as evidence that a stock is a "prudent" investment. Time variation in these imperfections can then induce time-varying clientele preferences. The 1970s witnessed a number of events that may have led to clientele demand shifts. The 1974 ERISA may have increased the attractiveness of payers to pension funds (Del Guercio (1996) and Brav and Heaton (1998)). The 1975 advent of negotiated commissions reduced the transaction cost of creating homemade dividends. And of course tax code changes can differentially affect payers and nonpayers.

Given that categorization occurs, time-varying demand between categories could also arise from what Mullainathan (2002) calls categorical inference. Investors using categorical inference could, for example, overestimate the impact of news about a particular payer for other payers, and underestimate its impact for nonpayers. Thus even without any explicit preference for cash dividends, the fact that categories have already been built around them could lead to variation in demand between payers and nonpayers.

Finally, building on ideas in Thaler and Shefrin (1981), Shefrin and Statman (1984) propose that some investors prefer dividend-paying stocks to homemade dividends to combat self-control problems. Shefrin and Statman also motivate an investor preference for dividends with prospect theory and regret aversion arguments. The prospect theory argument combines ideas in Kahneman and Tversky (1979) and Thaler (1980, 1983) with the result that dividends and capital gains allow investors a more flexible and agreeable mental accounting. When capital gains are low, investors can find a silver lining in the dividend; when capital gains are high, dividends and capital gains are individually-wrapped presents that can be savored separately. These theories offer additional reasons why investors view payers and nonpayers as distinct. To the extent that the germane considerations vary over time, they might also lead to a time-varying preference for payers.

B. Limited arbitrage

In perfect and efficient markets, uninformed demand for dividends would not affect stock prices. Arbitrage would prevent it. Arbitrageurs could short the firm with a preferred dividend policy and go long a correctly priced "perfect substitute" – a firm with the same investment policy but a different dividend policy. In perfect and efficient markets, only investment policy affects stock prices, so an arbitrage follows by making homemade dividends on the long firm to match the dividends declared by the short firm. In the absence of further frictions, this position delivers an up-front gain and can be risklessly held forever, or liquidated when prices move back in line. Competition for such arbitrage opportunities, it is argued, would eliminate any dividend premium or discount and maintain dividend policy irrelevance.

In practice, the long-short arbitrage that drives this irrelevance proof is risky and costly.[4] Limited arbitrage is the second postulate of the catering theory. An obvious risk in long-short arbitrage is fundamental risk, which arises simply because individual stocks do not have perfect substitutes (Wurgler and Zhuravskaya (2002)). This risk is in principle diversifiable, but arbitrageurs also face a systematic risk, often called noise-trader risk or interim price risk, if they try to trade against systematic sentiment. With short horizons or limited capital, they are sensitive to this risk (De Long, Shleifer, Summers, and Waldmann (1990) and Shleifer and Vishny (1997)). Finally, long-short arbitrage is costly. Nontrivial shorting costs are reported by D'Avolio (2002), Geczy, Musto, and Reed (2002), and Lamont and Jones (2002).

If arbitrage is limited and uninformed demand varies at the category level, as Barberis and Shleifer propose, then prices may also vary at the category level. Barberis, Shleifer, and Wurgler (2001) and Greenwood and Sosner (2001) find evidence for demand-induced comovement within the categories defined by stock indexes. If payers and nonpayers are

[4] Limited arbitrage explanations have been developed for closed-end fund discounts (Lee, Shleifer, and Thaler (1991) and Pontiff (1996)), risk arbitrage returns (Mitchell and Pulvino (2001) and Baker and Savasoglu (2002)), post-earnings-announcement drift (Mendenhall (2002)), the Internet bubble (Ofek and Richardson (2002a, 2002b)), seasoned equity issue returns (Pontiff and Schill (2001)), negative stub values (Lamont and Thaler (2000) and Mitchell, Pulvino, and Stafford (2001)), IPO underpricing (Duffie, Garleanu, and Pedersen (2002)), index inclusion effects (Greenwood (2001) and various papers on S&P 500 additions), and the predictive power of such variables as breadth of ownership (Chen, Hong, and Stein (2002)), market liquidity (Baker and Stein (2002)), and book-to-market (Alti, Hwang, and Trombley (2002)).

investment categories, the same logic implies that uninformed demand may also affect their relative prices.

Our empirical work is soon to come. For the impatient reader, we point to Long (1978) as some initial evidence that uninformed, time-varying demand for cash dividends affects stock prices. Long studies the Citizens Utilities Company, which between 1956 and 1989 had one share class that paid cash dividends and another that paid stock dividends. By charter, the payouts to the two classes were supposed to have equal pretax value. In practice, the stock dividend averaged ten percent higher than the cash dividend. Long finds that during his sample period, the cash-paying share's relative price was too high, given its pretax dividend disadvantage and its further tax disadvantage.[5] More interesting, the relative price fluctuates substantially over time. Long, Poterba (1986), and Hubbard and Michaely (1997) conclude that these fluctuations cannot be explained by traditional theories of dividends.

C. Catering as a rational response

The third element of the theory is that managers cater to uninformed investor demand. In the setting of dividends, catering implies that managers will tend to initiate or continue paying dividends when investors put a higher price on payers, and omit dividends or avoid initiating them when investors favor nonpayers.

The objective of catering is to capture the stock price premium associated with the characteristics investors currently favor. Catering is thus distinct from the usual policy of maximizing shareholder value. In inefficient markets, managers have to decide which of two prices to maximize: A short-run price affected by uninformed demand, and a fundamental or

[5] In 1955 CU obtained a special IRS exemption making the stock dividends not taxable as ordinary income. In general, regular stock dividends have been taxable since the 1969 Tax Reform Act, but CU received a grandfather clause in that Act.

long-run value determined by investment policy. Catering maximizes the short-run price, while the traditional policy emphasizes fundamental value.

In general, whether managers will rationally cater to a short-run mispricing is an empirical question.[6] One element in their decision is how much of a fundamental tradeoff there between catering and investment policy – if they can maximize short-run and long-run price without conflict, they will do both.[7] Another element is whether managers can personally profit from any short-term overvaluation that follows from successful catering. If they hold a significant amount of equity themselves, they can sell their overvalued shares. Or they may be able to issue dilutive, overpriced shares. A final consideration is the horizon of managers, or the horizon of the investors they care about most. Managers with short horizons, for instance those with compensation tied to short-run performance, will be more likely to cater.

D. A model of dividend catering

A short model makes these tradeoffs precise, and illustrates the more subtle features and limits of the catering theory. The model assumes that investors strictly categorize payers and nonpayers. While extreme, this is a convenient way of capturing the distinction that we want to emphasize – zero versus nonzero payout, not small versus large payout. Fama and French (2001) also focus on this dimension of dividend policy.

Consider a firm with Q shares outstanding. At $t = 1$, it pays a liquidating distribution of $V = F + e$ per share, where e is a normally distributed error term with mean zero. At $t = 0$, it has the

[6] Conditions under which managers will pursue short-run over long-run value are also discussed by Miller and Rock (1985), Stein (1989), Shleifer and Vishny (1990), Blanchard, Rhee and Summers (1993) and Stein (1996).

[7] An example of a setting in which no tradeoff exists is firm names. Cooper, Dimitrov, and Rau (2001) and Rau, Patel, Osobov, Khorana, and Cooper (2001) document that when investor sentiment favored the Internet (before March 2000), a number of firms added "dot com" to their names, but when sentiment turned away (after March 2000), firms were changing back. While many of these name changes surely coincided with changes in investment policy, Rau et al. argue that at least some of them were simply catering to sentiment for the Internet.

choice of paying an interim dividend $d \in \{0,1\}$ per share, which reduces the liquidating distribution by $d(1+c)$. The risk-free rate is zero. The cost c captures any tradeoff between dividend and investment policy, such as would result from costly external finance or taxes. The Miller and Modigliani case has c equal to zero – dividend policy does not in any way affect the cash flows to investors.

There are two types of investors, category investors and arbitrageurs. Both have constant absolute risk aversion. Arbitrageurs have aggregate risk tolerance per period of g^A. They have rational expectations over the terminal distribution, and they know the long-run cost of an interim dividend. Thus they expect a liquidating distribution of F if the firm does not pay an interim dividend and F-c if it does.

Category investors have aggregate risk tolerance per period of $g^C = g$. They have an irrational expectation of the terminal distribution, and they do not recognize the cost of an interim dividend. Their irrational expectation introduces a source of uninformed demand. For purposes of developing the model, we suppose that they categorize because they view nonpayers as growth firms, and they judge the prospects of those firms relative to their own assessment of growth opportunities. (Alternatively, their irrational expectations could reflect biased inferences that overweight within-category information as in Mullainathan (2002), biased risk perceptions arising from the bird-in-the-hand fallacy, or capture institutional constraints in a reduced form.) Specifically, they expect a final payment of V^D from payers and V^G from nonpayers. For simplicity, we assume that they misestimate the mean payout, but not the distribution around the mean. Typically, their net result is to cause V^D and V^G to fall on opposite sides of F.

If the firm meets its criteria, investor group k demands

$$D_0^k = g^k\left(E(V) - P_0\right). \tag{1}$$

Prices of dividend payers P^D (cum dividend) and growth firms P^G are therefore

$$P_0 = \begin{cases} P_0^D \equiv \frac{g}{g+g^A} V^D + \frac{g^A}{g+g^A}(F-c) - \frac{Q}{g+g^A} \\ P_0^G \equiv \frac{g}{g+g^A} V^G + \frac{g^A}{g+g^A} F - \frac{Q}{g+g^A} \end{cases} . \tag{2}$$

Given these prices, the manager chooses whether to pay dividends. We assume that the manager is risk neutral and cares about both the current stock price and the value of total distributions. The manager's only effect on the latter is through the cost of dividends c. With his horizon measured as l, the manager solves:

$$\max_d (1-l)P_0 + l(-dc) \tag{3}$$

The solution is straightforward. The manager pays dividends if the dividend premium is positive and exceeds the present value of the long-run cost that he incorporates. That is, when

$$P_0^D - P_0^G \equiv \frac{g}{g+g^A}\left(V^D - V^G\right) - \frac{g^A}{g+g^A} c \geq \left(\frac{l}{1-l}\right) c . \tag{4}$$

The first term in the middle is the immediate positive price impact of switching categories. The second is the immediate negative price impact of the arbitrageurs' recognition of the cost c. To induce payment, the net of these must exceed the long-run cost that the manager incorporates, the term on the right. Qualitatively, the propensity to pay is increasing in the dividend premium, decreasing in c, decreasing in the prevalence of arbitrage (the relative risk-bearing capacity of arbitrageurs and category investors), and decreasing in managers' horizons. The announcement effect of an initiation is positive and increasing in the dividend premium.[8]

[8] Note that an uninformed demand interpretation of announcement effects could explain why dividend changes have price impacts while at the same time appear to contain more information about past earnings than future earnings (Lintner (1956), Fama and Babiak (1968), Watts (1973), DeAngelo, DeAngelo, and Skinner (1996) and Benartzi, Michaely, and Thaler (1997)).

Equation (4) contains the basic time series predictions that we test, plus several cross-sectional predictions that we leave to future work. However, this two-category version is too simplistic to incorporate key stylized facts, such as the persistence of dividend payment and the negative announcement effect of omissions. To address these facts, we briefly outline extensions of the model that make use of a third category, *former* payers. These stocks lack the characteristics noticed by category investors, as they pay no dividends and have low (historical) earnings growth.[9] Thus they attract only arbitrageurs, so their price is $P_0^{FD} = F - \frac{Q}{g^A}$.

With this third category, the model can address the stylized fact that dividend payment is empirically quite persistent. That is, equation (4) shares the feature of many theories of dividends (for example, Miller and Rock (1985)) that the decisions to initiate and omit are symmetric. With former payers, dividends can be sticky. In particular, the decision for growth firms to initiate is still governed by (4), while current payers *continue* when:

$$P_0^D - P_0^{FD} \equiv \frac{g}{g+g^A}\left(V^D - \left(F - \frac{Q}{g^A}\right)\right) - \frac{g^A}{g+g^A}c \geq \left(\frac{l}{1-l}\right)c. \qquad (5)$$

Like the propensity to initiate, the propensity to continue is decreasing in the long-run cost and increasing in the dividend premium. The new insight is that continuing may be desirable even when initiating is not. More formally, if g^A is small, or if c is small *and* V^G and V^D fall on opposite sides of F, then (5) is satisfied whenever (4) is satisfied. Intuitively, former payers are neglected stocks, attracting only arbitrageurs. Even if initiating is undesirable, current payers may want to continue if the price impact to omitting is large. Note that this third category also suggests why

[9] The low historical earnings growth can be motivated by assuming that former payers' past dividends were not fully replenished by stock issues (perhaps as a result of the same external finance costs represented by c) or, more intuitively, on empirical grounds. Fama and French (2001) report that dividend payers have average (asset) growth rates of 8.78%, while nonpayers average 11.62% and former payers average only 4.67%. These averages are for the 1963-98 full sample. Between 1993-98, the averages are 6.65%, 17.67%, and 7.61% respectively.

some firms might initiate (reinitiate) dividends even when the dividend premium is negative, and why such initiations would still have a positive announcement effect.

A third category is also useful in addressing the stylized fact that the announcement effect of omissions is negative (Healy and Palepu (1988) and Michaely, Thaler, and Womack (1995)), whereas in the simplest two-category model it is not. Specifically, consider an intermediate time period between $t = 0$ and $t = 1$, in which the neglected former payers face a positive probability of being recategorized as growth firms – for example, because of a random earnings shock. In this setup, dividend payers may choose to omit a dividend at $t = 0$ even when (5) is not satisfied. They suffer a short-run negative announcement effect, but the expected value of being recategorized may be worth it. It is straightforward to formally incorporate this effect.

Of course, there are many other ways to explain some of these facts, such as fundamental risk, financial constraints, or asymmetric information. Our goal here is to illustrate the pros and cons of a model that isolates the market efficiency assumption of Miller and Modigliani. Such a model predicts that the propensity to pay dividends is robustly increasing in the dividend premium, and decreasing in the long-run costs of paying dividends. Realistic variants of it suggest that the decisions to initiate and to continue paying should be analyzed separately.

III. Empirical tests

We test the prediction that the decision to pay dividends depends on uninformed demand for dividend payers as revealed through stock price signals. The model illustrates some cross-sectional wrinkles, but this is primarily a time series prediction because uninformed demand is hypothesized to be systematic.

A. Dividend payment variables

Our measures of dividend payment are derived from aggregations of Compustat data. The observations in the underlying 1962-2000 sample are selected as in Fama and French (2001, p. 40-41): "The Compustat sample for calendar year *t* … includes those firms with fiscal year-ends in *t* that have the following data (Compustat data items in parentheses): total assets (6), stock price (199) and shares outstanding (25) at the end of the fiscal year, income before extraordinary items (18), interest expense (15), [cash] dividends per share by ex date (26), preferred dividends (19), and (a) preferred stock liquidating value (10), (b) preferred stock redemption value (56), or (c) preferred stock carrying value (130). Firms must also have (a) stockholder's equity (216), (b) liabilities (181), or (c) common equity (60) and preferred stock par value (130). Total assets must be available in years *t* and *t*-1. The other items must be available in *t*. … We exclude firms with book equity below $250,000 or assets below $500,000. To ensure that firms are publicly traded, the Compustat sample includes only firms with CRSP share codes of 10 or 11, and we use only the fiscal years a firm is in the CRSP database at its fiscal year-end. … We exclude utilities (SIC codes 4900-4949) and financial firms (SIC codes 6000-6999)."

Within this sample we count a firm-year observation as a dividend payer if it has positive dividends per share by the ex date, else it is a nonpayer. To aggregate this firm-level data into useful time series, two aggregate identities are helpful:

$$Payers_t = New\ Payers_t + Old\ Payers_t + List\ Payers_t\ , \quad (6)$$

$$Old\ Payers_t = Payers_{t-1} - New\ Nonpayers_t - Delist\ Payers_t\ . \quad (7)$$

The first identity defines the number of payers and the second describes the evolution. *Payers* is the total number of payers at time *t*, *New Payers* is the number of initiators among last year's nonpayers, *Old Payers* is the number of payers that also paid last year, *List Payers* is the number of firms that are payers this year and were not in the sample last year, *New Nonpayers* is the number of omitters among last year's payers, and *Delist Payers* is the number of last year's payers that are not in the sample this year. Note that two analogous identities hold if one switches "*Payers*" and "*Nonpayers*" everywhere. Also note that lists and delists are with respect to our sample, which involves several screens. Thus new lists include both IPOs that survive the screens in their Compustat debut as well as established Compustat firms when they first survive the screens. It also includes the established NASDAQ firms that appear in Compustat for the first time in the 1970s. Similarly, delists include both delists from Compustat and firms that fall below the screens.

We use these aggregate totals to define three basic measures of the dynamics of dividend payment among certain subsets of firms:

$$Initiate_t = \frac{New\ Payers_t}{Nonpayers_{t-1} - Delist\ Nonpayers_t}, \tag{8}$$

$$Continue_t = \frac{Old\ Payers_t}{Payers_{t-1} - Delist\ Payers_t}, \tag{9}$$

$$Listpay_t = \frac{List\ Payers_t}{List\ Payers_t + List\ Nonpayers_t}. \tag{10}$$

In words, the rate of initiation *Initiate* is the fraction of surviving nonpayers that become new payers. The rate at which firms continue paying *Continue* is the fraction of surviving payers that continue paying. It can also be viewed as one minus the rate at which firms *omit* dividends. The rate at which new lists in the sample pay *Listpay* is self-explanatory.

These variables capture the decision whether to pay dividends, not how much to pay. We take this approach for several reasons. First, these are the natural dependent variables in a theory in which investors categorize shares based on whether they pay dividends. (Wings make a "bird," regardless of their length.) Second, as an empirical matter, the payout ratio is sensitive to profitability and the dividend yield is sensitive to changes in share prices. The decision to initiate or omit dividends, in contrast, is always a policy decision. Third, Fama and French (2001) document a decline in the number of payers, and no comparable pattern in the payout ratio. Nonetheless, measures of the level of dividends turn out to be useful in discriminating among alternative interpretations for the basic results.

Table 1 lists the aggregate totals and the dividend payment variables. The sample displays similar characteristics to the sample in Fama and French (2001). For our purposes, the most notable feature of the data is the time variation in the dividend variables. The rate of initiation starts out high in the early years of the sample, then drops dramatically in the late 1960s, rebounds in the mid 1970s, drops again in the late 1970s and remains low through the end of the sample. The rate at which firms continue paying displays less variation, as expected. The rate at which lists pay displays the most variation. As Fama and French point out, it has declined steadily over the past few decades.

While we do not focus on the level of dividends, as just discussed, it is useful to get a rough sense of the aggregate economic significance of initiations. In the average year in our sample, newly-initiated dividends amount to 0.5% of dividends already paid by payers, and 29% of the change in the amount that is paid by payers (in years when this change is positive). The fact that the first number is so small is not surprising. The numerator is small because the rate of initiation is low and the typical initiator is small and starts off with a small dividend, while the denominator is high because the persistence of payment is high and the typical surviving payer tends to increase dividends over time. We also caution that the 29% figure is affected by outlying years in which the change in the amount paid by existing payers is barely positive. Nonetheless, these figures provide some sense of the aggregate economic significance of initiations.

B. Investor demand for dividends variables

We relate dividend payment choices to several stock market measures of the uninformed demand for dividend-paying shares. Conceptually, an ideal measure would be the difference between the market prices of firms that have the same investment policy and different dividend policies. In the frictionless and efficient markets of Miller and Modigliani (1961), of course, this price difference is zero. But uninformed demand combined with limits to arbitrage, as discussed above, can lead to a time-varying price difference.

Our first measure, which we simply call the dividend premium because it is the broadest measure, is motivated by this intuition. It is the difference in the logs of the average market-to-book ratios of payers and nonpayers – that is, the log of the ratio of average market-to-books.[10] We define market-to-book following Fama and French (2001). Market equity is end of calendar year stock price times shares outstanding (Compustat item 24 times item 25).[11] Book equity is stockholders' equity (Item 216) [or first available of common equity (60) plus preferred stock par value (130) or book assets (6) minus liabilities (181)] minus preferred stock liquidating value (10) [or first available of redemption value (56) or par value (130)] plus balance sheet deferred taxes and investment tax credit (35) if available and minus post retirement assets (330) if available. The market-to-book ratio is book assets minus book equity plus market equity all divided by book assets.

[10] Market-to-book ratios are approximately lognormally distributed. As a result, levels of the market-to-book ratio, unlike logs, have the property that the cross-sectional variance increases with the mean. In our context, this means that the absolute size of a premium measured in levels could proxy for a market-wide valuation ratio.

[11] Here we want an aggregate market-to-book measure for a precise point in time, the end of the calendar year. Later in the paper, when we use market-to-book as a firm characteristic, we use the end of fiscal year stock price.

We then average the market-to-book ratios across payers and nonpayers in each year. The equal- and value-weighted dividend premium series are the difference of the logs of these averages. These variables are listed by year in Table 2 and the value-weighted series are plotted in Figure 1. The figure shows that the average payer and nonpayer market-to-books diverge significantly at short frequencies. It reveals several interesting patterns. Dividend payers start out at a premium, by this measure, in the first years of the sample. The valuation of nonpayers then spikes up in 1967 and 1968 and falls sharply, in relative terms, through 1972. The dividend premium takes another dip in 1974, and for over two decades now payers have traded at a discount by this measure. The discount widened in 1999 but closed somewhat in 2000. At this point it is premature to speculate on the forces that move the dividend premium variable over time. In Baker and Wurgler (2002b), we draw on academic histories of the capital market and a review of historical articles in the financial press to provide a detailed, but still highly stylized, account of its variation.

The primary disadvantage of the dividend premium variable is that it may also reflect the relative investment opportunities of payers and nonpayers, as opposed to uninformed demand for dividend-paying shares. We consider this in our discussion of alternative explanations.

Our second measure of investor demand for dividend payers is the difference in the prices of Citizens Utilities cash dividend and stock dividend share classes. Between 1956 and 1989 the Citizens Utilities Company had two classes of shares outstanding on which the payouts were to be of equal value, as set down in an amendment to the corporate charter. In practice, the relative payouts were close to a fixed multiple. Long (1978) describes the case in detail. We measure the CU dividend premium as the difference in the log price of the cash payout share and the log price of the stock payout share. The 1962 through 1972 data were kindly provided by John Long and the 1973 through 1989 data are from Hubbard and Michaely (1997).[12] Table 3 reports the CU dividend premium year by year.

By its nature, the CU dividend premium does not reflect anything about investment opportunities. This reduces the number of alternative explanations for why it fluctuates, but it also means that arbitraging the premium entails no fundamental risk, only noise-trader risk, so the amount of sentiment that it reflects may be muted. Other disadvantages include the fact that CU is just one firm; the stock payout share is more liquid than the cash payout share; there was a one-way, one-for-one convertibility of the stock payout class to the cash payout class, truncating the ability of the price ratio to reveal pro-cash-dividend sentiment; certain sentiment-based mechanisms outlined above involve categorization of firms rather than shares, so a case in which one firm offers two dividend policies may lead to weaker results; and the experiment ended in 1990, when CU switched to stock payouts on both classes.

Our third measure of uninformed demand for dividends is the average announcement effect of recent initiations.[13] Intuitively, if investors are clamoring for dividends, they may make themselves heard through their reaction to initiations. Asquith and Mullins (1983) find that initiations are greeted with a positive return on average, but they do not study whether this effect varies over time. We define a dividend initiation as the first cash dividend declaration date in CRSP in the twelve months prior to the year in which the firm is identified as a Compustat New Payer. Since Compustat payers are defined using fiscal years while CRSP allows us to use

[12] There are two further adjustments made throughout the 1962 through 1989 series. The annual value that we consider is the log of the average of the monthly price ratios, because the relative prices fluctuate dramatically even within a year. And to control for the fact that cash dividends were quarterly, in practice, while the stock dividends were semiannual, the cash dividends are assumed to be reinvested until the corresponding stock dividend is paid.

[13] In closer analogy with the other demand variables, one might like to define an announcement effect variable that combines the reactions to initiations and omissions. That is, when demand for dividend payers is high, initiation effects may be particularly positive and omission effects particularly negative. Unfortunately, CRSP data do not provide precise omission announcement dates.

calendar years, the resulting asynchronicity means that the number of initiation announcements identified in CRSP for year *t* does not equal the number of Compustat New Payers in year *t*. Another difference arises because the required CRSP data are not always available.

Given an initiation in calendar year *t*, we calculate the cumulative abnormal return over the three-day window from day –1 to day +1 relative to the CRSP declaration date as the cumulative difference between the firm return and the CRSP value-weighted market index. To control for the differences in volatility across firms and time (see Campbell, Lettau, Malkiel and Xu (2000)), we scale each firm's three-day excess return by the square root of three times the standard deviation of its daily excess returns. The standard deviation of excess returns is measured from 120 calendar days through five trading days before the declaration date. Averaging these across initiations in year *t* gives a standardized, cumulative abnormal announcement return *A*. To determine whether the average return in a given year is statistically significant, we compute a test statistic by multiplying *A* by the square root of the number of initiations in year *t*. This statistic is asymptotically standard normal and has more power if the true abnormal return is constant across securities (Brown and Warner (1980) and Campbell, Lo, and MacKinlay (1997)), which is a natural hypothesis in our context. Table 3 reports the average standardized initiation announcement effects year by year.

Our last demand measure is the difference between the *future* returns on value-weighted indexes of payers and nonpayers. Under the rather stark version of catering outlined in the previous section, managers rationally initiate dividends to exploit a market mispricing. If this is literally the case, then a high rate of initiations should forecast low returns on payers relative to nonpayers as the overpricing of payers reverses. The opposite should hold for omissions.

Table 4 reports the correlation between the sentiment measures. We correlate the first three measures at year t with the excess real return on payers over nonpayers r_D - r_{ND} in year t+1 and the cumulative excess return R_D - R_{ND} from years t+1 through t+3. To the extent that these variables capture a common factor in uninformed investor demand for dividends, we expect the dividend premium, the CU premium, and announcement effects to be positively correlated with each other, and negatively correlated with the future excess returns of payers. Table 4 shows that these correlations are as expected, with two exceptions: the CU premium and the initiation effect are negatively correlated, and the initiation effect and one-year-ahead excess returns are positively correlated. The dividend premium is correlated with all of the other variables in the expected direction, however. This suggests that the dividend premium may be the single best reflection of the common factor. In any case, given that each measure has its own advantages and disadvantages, it is reassuring that they correlate roughly as expected.[14]

Table 4 also reports autocorrelations and Dickey-Fuller tests for unit roots. These statistics shed light on the time series properties of the data and the potential for spurious correlation in the regressions to follow. Of course, the textbook case of spurious correlation involves nonstationary variables, and so before one puts too much weight on the Dickey-Fuller tests it is worth noting the theoretical considerations that suggest that these variables are indeed stationary. For example, if the market-to-book ratio is itself stationary, the dividend premium cannot grow without bound. In the absence of this prior information, however, Table 4 shows that we cannot reject a unit root in the dividend premium or the CU dividend premium. A similar logic holds for the dividend payment variables: Each one is mathematically bounded between

[14] We have also considered average ex-dividend day returns as a fifth measure of investor demand. Ex-day returns do vary over time (e.g., Eades, Hess, and Kim (1994)). However, they have less of a category-switching interpretation than our other four measures: A dividend payer seems likely to be viewed as a payer before, during, and after the ex-day.

one and zero, but we cannot formally reject a unit root (unreported). More practically, what these statistics suggest is that in certain cases we should control for a time trend before concluding that a relationship is robust.

C. Time series relationships

Here we document the basic relationships between the rates of dividend payment and the measures of the demand for dividend-paying shares. The top panel of Figure 2 plots the dividend premium against the raw rate of dividend initiation in the following year.

The figure reveals a strong positive relationship, consistent with catering. On average, the rate of initiation is 11.0% when the dividend premium is positive and only 3.1% when it is negative. In the first half of the sample, the dividend premium and subsequent initiations move almost in lockstep. The premium then submerges in the late 1970s, leading the rate of initiation down once again. A qualitatively similar figure obtains with the rate of initiations by large firms, small firms, or firms that have been listed for at least five years (unreported).

The dividend premium has been negative since around 1978, and the initiation rate has also remained low. The figure gives a visual impression that the relationship has broken down in this period. In fact, this pattern is not inconsistent with the theory. Equation (4) indicates that there is no reason to initiate dividends when they are discounted. A monotonic relationship between initiations and the dividend premium is predicted only when the latter is positive. Consistent with this prediction, the correlation between the two series is 0.53 for the 14 years in which the lagged dividend premium is positive, and 0.03 for the 24 years in which it is negative. Of course, another (less flattering) possibility is that exogenous factors such as the growth in dividend-unprotected executive stock options, or the emergence of repurchases as a substitute for dividends, have suppressed initiations in recent years. Still another interpretation of the 1980s and 1990s data is mentioned below (where we discuss the lower panel of Figure 2).

To examine the basic relationship in the figure more formally, Table 5 regresses the dividend policy measures on the lagged demand for dividends measures. For example, the initiation rate is modeled as:

$$Initiate_t = a + bP_{t-1}^{D-ND} + cA_{t-1} + dP_{t-1}^{CU} + u_t \, , \tag{11}$$

where *Initiate* is the rate of initiation, $P^{D\text{-}ND}$ is the market dividend premium (value-weighted or equal-weighted), A is the average initiation announcement effect, and P^{CU} is the Citizens Utilities dividend premium. All independent variables are standardized to have unit variance and all standard errors are robust to heteroskedasticity and serial correlation to four lags using the procedure of Newey and West (1987).

Panel A reports the determinants of initiations. The regression in the first column corresponds to Figure 2. It shows that a one-standard-deviation increase in the value-weighted market dividend premium is associated with a 3.90 percentage point increase in the initiation rate in the following year, or roughly three-quarters of the standard deviation of that variable.[15] This one measure explains a striking 60 percent of the variation in the rate of initiation. The second column shows that the effect of the equal-weighted dividend premium is essentially the same.[16] The remaining columns show the effect of other variables, and the results of a multivariate horse

---

[15] If nonpayers are trading at a discount to payers, a large number of initiations may mechanically dilute the price of payers and hence lower the premium. This can create the sort of Stambaugh (1999) bias that is described in the Appendix in connection with return predictability. This bias is increasing in the correlation between the errors of the prediction regression in Table 5 and the errors in an autogression of the dividend premium on the lagged dividend premium. In the case of *Initiate*, these errors have a correlation of less than 0.01, so the bias is inconsequential. In the case of *Continue* and *Listpay*, the correlation is also not statistically significant.

[16] The dependent variable is implicitly an equal-weighted measure, so an equal-weighted independent variable may seem appropriate. On the other hand, the value-weighted premium, which emphasizes larger firms, is likely to be more visible to potential initiators. The two measures perform almost identically in this and future tables. We proceed with value weights alone for the sake of brevity.

race. The lagged initiation announcement effect and the CU premium have significant positive coefficients, as predicted. But they disappear in a multivariate regression that includes the dividend premium. This is consistent with earlier indications that the dividend premium best captures the common factor in these variables.

Panel B reports analogous regressions for the rate of continuation. The dividend premium effect is again as predicted by catering: When dividends are at a *discount*, payers are more likely to *omit* (not continue). The dividend premium effect is smaller here, consistent with the lower sensitivity predicted by certain versions of the model. Specifically, a one-standard-deviation increase in the dividend premium increases the continuation rate by 0.85 percentage points. Indeed, to the extent that some omissions are forced by profitability circumstances, which we control for in the next section, it may be surprising that the effect is as strong as it is. The other columns of Panel B show that the other measures of demand do not have explanatory power for the rate of continuation, however.

Panel C shows that the rate at which lists are pay is also positively related to the dividend premium. A one-standard-deviation increase in the dividend premium increases *Listpay* by 16.08 percentage points. The relative size of the coefficient here again reflects the relative variation in the dependent variable. Using a dividend premium variable defined just over recent new lists has at least as much explanatory power (unreported). The CU premium also has a strong univariate effect here, but as before the dividend premium wins a horse race.

Table 6 shows the relationship between dividend policy and our fourth demand variable, the future excess returns of payers over nonpayers. In Panel A, the dependent variable is the difference between the returns on value-weighted indexes of payers and nonpayers. Panels B and C look separately at the returns on payers and nonpayers, respectively, to examine whether any results for relative returns are indeed coming from the difference in returns, which the theory emphasizes, and not payer or nonpayer returns alone. Each panel examines one, two, and three-year ahead returns, and cumulative three-year returns. The table reports ordinary least-squares coefficients as well as coefficients adjusted for the small-sample bias analyzed by Stambaugh (1999). The p-values reported in the table represent a two-tailed test of the hypothesis of no predictability using a bootstrap technique described in the Appendix.

Panel A indicates that dividend policy does have predictive power for relative returns. A one-standard-deviation increase in the rate of initiation forecasts a decrease in the relative return of payers of around eight percentage points in the next year, and thirty percentage points over the next three years. This strikes us as a substantial magnitude – arguably, a magnitude worth catering to. The predictive power of the standardized continuation rate is similar. The rate at which lists pay has no predictive power, however, unless a time trend is included, in which case it displays a similar level of predictability to the other dividend policy variables. The bottom panels confirm that the relative return predictability cannot be attributed to just payer or nonpayer predictability. As theory suggests, it is the relative return that matters.

Tables 5 and 6 provide some support for the catering theory's basic predictions. Firms appear more likely to initiate dividends when the demand for dividend-paying shares is high, and more likely to omit when demand is low.

IV. Alternative explanations

The catering explanation for these results is that dividend payment is, to an important extent, a rational managerial response to a real or perceived stock market mispricing. While it is often possible to reinterpret an individual empirical relationship, it turns out to be very difficult to construct a coherent, non-catering alternative explanation for the full set of results. We discuss a variety of alternative hypotheses below.

A. Statistical robustness

Time-series regressions raise a standard set of statistical issues. One is spurious correlation. Recall that despite theoretical reasons to believe that the variables are stationary, formal tests do not always reject a unit root. On the other hand, the key for statistical inference is that the residuals are stationary. In the regression of initiation on the dividend premium, we can reject a unit root in the residuals at the 10 percent level (unreported).

A more practical question is whether these relationships are robust to the inclusion of a time trend. One would not expect future relative stock returns to be predictable from a time trend, but the other measures of investor demand are worth checking. Table 7 includes a trend alongside the dividend premium. The coefficient remains strongly significant for initiations. For continuations, however, inclusion of a trend pushes the coefficient to the 10 percent level of significance, and considerably reduces the size of the coefficient on new lists, though it does not eliminate statistical significance.

In unreported results, we include a trend alongside the CU dividend premium and the initiation announcement effect. This changes our earlier inferences only in the case of the CU dividend premium: It does not have explanatory power beyond a common trend. We have also considered the raw (unstandardized) average initiation announcement effect, which we did not examine earlier. It turns out to have a positive but insignificant univariate relationship with initiations; however, it becomes significant in the presence of a trend term.

B. Time-varying investment opportunities

We now turn to economics-based alternative explanations. The relationship in Figure 2 could be an artifact of time variation in investment opportunities, in an environment of rational managers and rational investors. That is, nonpayers may be initiating dividends not because they are chasing the relative premium on payers but because their investment opportunities are low in an absolute sense. An inverse relationship between dividends and investment opportunities could follow if external finance is costly, as in Myers (1984) and Myers and Majluf (1984), or if dividends are a response to agency costs of free cash flow, as in Jensen (1986). This is a natural alternative explanation that is worth considering in some detail.

A first point is that this explanation makes the converse prediction that payers will be more likely to *omit* when their investment opportunities are *high.* This would imply a negative relationship between the dividend premium and the rate at which firms continue paying, not positive as we found earlier. Therefore, this alternative could apply only to initiations.

To examine its bite (for initiations), a straightforward test is to simply control for the *level* of investment opportunities and see if the dividend *premium* retains residual explanatory power. We consider two potential measures of investment opportunities, the average market-to-book of the set of firms in question and the overall CRSP value-weighted dividend yield. The first and fourth columns in Table 7 show the results. The investment opportunities proxies enter with the predicted signs – nonpayers are less likely to initiate when their average market-to-book is high, and when the overall dividend-price ratio is low. For continuations and new lists, however, these variables enter with the wrong sign for the alternative explanation. More important, the dividend premium coefficient is not much affected.

The investment opportunities view also makes similar predictions for both repurchases and dividends, while catering involves only the latter. Thus we can examine whether the rate of repurchase is also related to the dividend premium, or only the rate of dividend initiation. We construct aggregate time series measures of the rate of repurchase, defining a repurchase as nonzero purchase of common and preferred stock (Compustat item 115). The first useable year is 1972. We find that the rate of repurchase among all firms, and the rate at which firms "initiate" repurchases (new repurchasers in year *t* divided by surviving non-repurchasers), have an insignificant negative correlation with the lagged dividend premium (unreported). The dividend initiation rate, by contrast, has a correlation of 0.73 over the same 29-year period.

Finally, time-varying investment opportunities leads more naturally to variation in the level of dividends, not necessarily the number of firms paying a dividend as is the essence of initiations and omissions. Thus, this alternative hypothesis would predict that the dividend premium should bear an even stronger relationship to the level of dividends, whereas catering to category investors would not necessarily predict a relationship in levels. Consistent with the latter view, we find that neither the payout ratio nor the dividend yield is significantly correlated with the lagged dividend premium (unreported), where we use updated data from Shiller (1989) on earnings and dividends for the S&P 500 and the CRSP value-weighted dividend yield. Also note that we control for the dividend yield directly in the last three columns of Table 7. This actually increases the effect of the dividend premium on the initiation rate.

All of this casts doubt on the ability of investment opportunities to explain the connection between initiations and the dividend premium. Moreover, this story has fundamental difficulties addressing the connection to future relative returns or the CU dividend premium (though as mentioned above, this relationship is not apparent beyond a common time trend).

C. Correlated errors in forecasting investment opportunities

The second alternative explanation we consider is a variant of the first and is suggested by the referee. Perhaps managers and investors make correlated errors in their forecasts. That is, investors sometimes get excited about growth prospects and bid up the price of nonpayers, who they feel are better suited to exploit new opportunities. Managers, rather than catering to this sentiment, are equally smitten and choose to invest all available resources rather than paying any dividends. This story is better than the rational expectations version outlined above in that it can address the return predictability results, but otherwise it has the same shortcomings.

D. Time-varying sample characteristics

The results could arise because our dividend demand measures are somehow related to the cross-sectional *distribution* of dividend-relevant characteristics *within* payer and nonpayer samples. This is a more general version of the investment opportunities explanation discussed above. As a contrived example, suppose the variance of investment opportunities among nonpayers increases (for some reason) whenever the dividend premium increases. Then an increasing initiation rate could indicate that a relatively high fraction of nonpayers do not need to retain cash, not that nonpayers as a group are catering to the dividend premium. Note that in this example, the average investment opportunities of nonpayers are held constant, so the time series exercises in Table 7 would mistakenly attribute the effect to the dividend premium.

We can evaluate this explanation by controlling directly for sample characteristics. In particular, we examine whether the dividend premium helps to explain the residual variation in dividend decisions after controlling for the characteristics studied by Fama and French (2001). They model the expected probability that a firm is a payer as a function of four variables:

$$\Pr\left(Payer_{it} = 1\right) = \text{logit}\left(a + bNYP_{it} + c\frac{M}{B}_{it} + d\frac{dA}{A}_{it} + e\frac{E}{A}_{it}\right) + u_{it}, \qquad (12)$$

where size *NYP* is the NYSE market capitalization percentile, i.e. the percentage of firms on the NYSE having smaller capitalization than firm *i* in that year. Market-to-book *M/B* is measured as defined previously, with the slight modification that here we use the fiscal year closing stock price (Compustat item 199) instead of the calendar year close. Growth *dA/A* in book assets (Compustat item 6) is self-explanatory. Profitability *E/A* is earnings before extraordinary items (18) plus interest expense (15) plus income statement deferred taxes (50) divided by book assets. The error term *u* is the residual propensity to pay dividends for a given firm-year.

The tests proceed in two stages. In the first stage, we follow Fama and French in estimating firm-level logit regressions using these firm characteristics. As before, we examine dividend payment separately among surviving nonpayers, surviving payers, and new lists. We also follow them in estimating specifications that exclude *M/B* – they suggest that the degree to which this variable measures investment opportunities may change over time, and indeed we have been arguing that this variable is affected by uninformed investor demand.

In the second stage, we regress the average annual prediction errors, or the aggregate "propensity to pay," on the value-weighted dividend premium. For example, naming *PTI* the residual rate of initiation or the "propensity to initiate," we estimate:

$$P\tilde{T}I_t = f + gP_{t-1}^{D-ND} + v_t \text{, where} \qquad (13)$$

$$P\tilde{T}I_t \equiv \tfrac{1}{N}\sum_i u_{it} \,.$$

Explanatory power for the propensity to initiate (or, analogously, the propensity to continue $P\tilde{T}C$ or propensity to list as a payer $P\tilde{T}L$) would mean that the dividend premium is not affecting dividend policy through the average *or* the cross-sectional distribution of any of these four characteristics.[17] The regression in (13) is analogous to our earlier time series regressions, such as equation (11), but now the effect of varying characteristics has been removed. Note that the two-stage approach gives deference to the characteristics variables by allowing the dividend premium to explain only residual variation. And in terms of statistical power, the dividend premium is using only 38 data points to fit, not thousands like the characteristics.

Table 8 shows the results of this exercise. The first stage results indicate that size and profitability have the most robust effects on the propensity to pay, as Fama and French find. The right column shows the second stage results. In general, controlling for characteristics directly, the dividend premium retains statistically significant explanatory power for most subsamples. Comparing these coefficients to our earlier time series results, one can see that controlling for firm characteristics barely affects the initiation rate coefficient. It is 3.90 in Table 5, and controlling for characteristics moves it only slightly, and does not affect its statistical significance. We view this as compelling evidence that the dividend premium is not working through a background correlation with the level *or* distribution of characteristics.

Controlling for characteristics also tends to improve the post-1980 correlation between initiations and the dividend premium, as shown in Panel B of Figure 2. This suggests another perspective on the poor post-1980 fit of the raw initiation rate. Namely, that the raw rate was depressed by the recent influx of small, unprofitable, high market-to-book firms noted by Fama and French. Within the language of the model, firms with these characteristics would tend to have high fundamental costs of paying dividends. Controlling for characteristics may better

[17] Including the dividend premium directly in equation (13) and estimating the coefficients in a panel regression gives qualitatively similar results to our two-stage procedure (unreported). A panel regression is necessary in that specification because the dividend premium does not vary within a year, as the Fama-MacBeth procedure requires.

reveal the partial effect of the dividend premium.[18] Interestingly, the only period where the rate of initiation is sharply lower than would be expected from the dividend premium is the early 1970s. The most obvious explanation is the Nixon dividend controls (1971-1974).

Controlling for characteristics does tend to reduce the effect of the dividend premium among the other samples, however. That characteristics would help to explain omissions might be expected given that they are known to be associated with characteristics such as low profitability. Nevertheless, the dividend premium approaches statistical significance even in this sample, and remains statistically significant in the new list sample.

The methodology in Table 8 is also useful for confirming once again that our empirical results, like our theory, are mainly about the decision *whether* to pay dividends, not *how much* to pay. That is, we have constructed a time series of the raw rate of dividend *increases* and found that it has a significant positive correlation with the dividend premium (unreported), but that this result comes entirely from changing characteristics like profitability. When these characteristics are accounted for using the two-stage procedure, there is no relationship between the residual propensity to increase dividends and the dividend premium (unreported).

Finally, we can also ask whether the average annual prediction errors from Table 8 predict the relative returns of payers and nonpayers. In other words, whether the non-characteristics-related variation in dividend policy, which is presumably more closely related to catering, also predicts returns. We find that the average prediction errors indeed have comparable

[18] To be fully consistent with the theory, this interpretation of the 1980s and 1990s also involves low-frequency measurement error in the dividend premium. If the true dividend premium was negative, then as mentioned before, the low post-1980 correlation with the raw level of initiations is not problematic for the theory. The fact that controlling for characteristics improves the correlation suggests that the true dividend premium may have been positive over some of this period, but that we measure it with low-frequency error. For example, if nonpayers typically have greater intangible assets, their market-to-books could be naturally higher than those of payers, but higher-frequency variation in the dividend premium might still be informative about demand for dividends.

or greater predictive power than the raw dividend payment measures (unreported). This indicates that the predictability results also do not work through firm characteristics.

E. Time-varying contracting problems

Another class of alternative explanations involves time-varying contracting problems, such as adverse selection or agency. With regard to adverse selection, it is possible that when nonpayers trade at a low value, this is a particularly important time for them to signal their investment opportunities. Initiating dividends serves as a signal in the models of Bhattacharya (1979), Hakansson (1982), John and Williams (1985), and Miller and Rock (1985). Once again, the natural way to evaluate this hypothesis is to control for the level of nonpayer market-to-book directly, and examine whether the dividend premium has residual explanatory power. The results in Table 7 show that it does. Moreover, it is hard to imagine a rational expectations equilibrium model in which dividend policy choices predict future returns, or would have any natural reason to be correlated with the CU dividend premium.

Agency costs may also vary over time, with high agency costs requiring dividend payments. For example, La Porta, Lopez-de-Silanes, Shleifer, and Vishny (2000) find that dividend policy varies across countries according to the degree of investor protection. If the dividend premium were a simple time trend, this could be a more compelling explanation for our results. As it stands, this explanation requires governance to improve briefly in the late 1960s, deteriorate, and then improve again. Of course, it is possible that variation in investment opportunities and profits might affect agency costs, but this would be addressed in Table 8. Here, one must imagine agency problems that arise independent of firm characteristics.

V. On the source of demand for dividends

Process of elimination, as well as the close connection between the results and predictions of the model, suggests that managers are catering to investor demand. Here we ask the follow-up question: Which investors are managers catering to? Put differently, what drives the dividend premium? These are hard questions and we offer only preliminary conclusions. The two basic possibilities are traditional dividend clienteles or sentimental investors.

A. Dividend clienteles

Black and Scholes (1974) suggest that uninformed demand for dividends result from dividend clienteles, which in turn derive from such imperfections as taxes, transaction costs, or institutional investment constraints.[19] In general, rational clienteles would be satisfied by a supply response in the aggregate level of dividends, not the number of dividend-paying shares. Also, if they are diversified, rational clienteles will not care about how the supply response is distributed across firms. In fact, Marsh and Merton (1987) point out that current dividend payers, with high financial slack and modest investment opportunities, are probably the lowest marginal cost suppliers of dividends. These considerations suggest that if the dividend premium were varying in response to rational clientele demands, it should have a closer connection to the level of dividends than the number of payers. We find the opposite.

Another approach is to see if we can directly match up the dividend premium with any plausible proxies for clienteles. A natural proxy for tax clienteles, for example, is the relative tax advantage of dividend income versus capital gains. Figure 1 suggests that the 1986 Tax Reform Act, which should have shrunk the anti-dividend clientele, had no visible effect on the dividend

[19] Miller and Scholes (1978) propose that tax code changes could have no influence, because taxes on dividends can be postponed indefinitely. However, Peterson, Peterson, and Ang (1985) find empirically that most investors do not avoid taxation.

premium. Similarly, Hubbard and Michaely (1997) study the reaction of the CU dividend premium to the 1986 reform. They conclude that tax-motivated clienteles do not seem to affect that variable. As an aside, the lack of a differential reaction to the reform by payers and nonpayers also seems inconsistent with dividend tax capitalization.

Table 7 contains a more formal test of whether firms are catering to tax clienteles. The personal tax advantage for dividends (typically a net disadvantage) is measured as the ratio of the after-tax income from a dollar of dividends to a dollar of long-term capital gains. That is, we take one minus the average marginal income rate, divided by one minus the average marginal long-term capital gains rate. The tax rates in this calculation are weighted average rates across shareholder groups as calculated by the NBER TAXSIM model. They are reported at www.nber.org/~taxsim/mrates/mrates2.html and described by Feenberg and Coutts (1993). Table 7 shows that if anything, the initiation rate is positively related to this variable, not negatively related, and in any case its inclusion does not much affect the dividend premium coefficient. (In Panel C, the large t-statistic on taxes disappears when a trend is included because of trends in both the rate at which lists pay and the tax advantage variable.)

Transaction costs also vary over time, changing the cost of homemade dividends. Perhaps this induces changes in demand by transaction cost clienteles. Black (1976) dismisses this argument, pointing out that there are simple institutional solutions to the problem of the small investor's transaction costs. However, Jones (2001) shows that transaction costs have declined dramatically since the mid-1970s, which coincides with the reduction in the rate of initiation that we document.[20] Jones's Figure 3 shows the average annual one-way transaction cost for the NYSE, or one half of the bid-ask spread plus commissions. This series is strongly positively

[20] The rise of mutual funds roughly coincides with these falling transaction costs, potentially lowering an individual investor's cost of monetizing capital gains further still.

correlated with the rate of initiation, though this comes mostly from a common time trend; the correlation between the detrended variables is not significant (unreported). More importantly, in regressions that include both variables, the dividend premium has more statistical significance than transaction costs in explaining the initiation rate (unreported).

Another theoretical possibility is that dividend clienteles are motivated by institutional investment constraints. For instance, the 1974 Employee Retirement Income Security Act may have increased the pro-dividend clientele by creating a vague "prudent man" rule for pension funds. The law was revised in 1979 to allow pension funds to provide venture capital, thus erasing any doubt that nonpayers were acceptable investments and perhaps shrinking the dividend clientele. Figure 2 could be broadly consistent with these institutional shifts. However, the dividend premium seems to anticipate the law, peaking in 1972 and beginning its drop in 1977. Perhaps ERISA is part of the story in this period, but we are not aware of investment constraints that could explain the dividend premium over the 1960s and early 1970s.

The rational clientele explanations for the dividend premium also face some difficulty accounting for the magnitude of the return predictability effects. Even under limited arbitrage, in equilibrium the marginal clientele investor should still be indifferent to leaving the clientele or taking advantage of the mispricing that his colleagues presumably induce. But the marginal clientele investor's savings on transaction costs or taxes, for example, seem unlikely to be worth a tradeoff of nine percentage points per year in pre-tax expected returns.

B. Sentiment for dividends

For these reasons we conclude that rational dividend clienteles are unlikely to be driving the dividend premium. This leaves time-varying sentiment between payers and nonpayers as the remaining explanation. Of course, economists are just beginning to understand sentiment, so such hypotheses are harder to reject by construction. Here we attempt to provide some rejectable tests for the sentiment-based view of demand for dividends.

We outlined two specific sentiment mechanisms earlier in the paper. One was based on the bird-in-the-hand fallacy and time-varying risk aversion. It proposes that when investors are highly tolerant of risk, they stray more from the perceived safety of dividend-paying stocks. Another story involves time-varying investor perceptions of growth opportunities. It holds that uninformed or unsophisticated investors use dividend policy to infer a firm's investment plans. From a zero-payout policy (controlling for profitability), they tend to infer that the firm wants to reinvest and grow. And so when they believe the outlook for growth stocks is generally good, they favor nonpayers. When they feel it is bad, they favor payers. Either sentiment mechanism seems consistent with most of our results, including return predictability.

As a first test of these mechanisms, we compare the closed-end fund discount with the dividend premium. Zweig (1973) and Lee, Shleifer, and Thaler (1991) view the closed-end fund discount as a measure of general investor sentiment. Whether it reflects risk tolerance, expectations for growth stocks, or both, is far from clear. A positive correlation between the closed-end fund discount and the dividend premium would therefore be consistent with both mechanisms outlined above, and would not be predicted by any of the alternative explanations we have considered. We gather value-weighted discounts on closed-end stock funds for 1962 through 1993 from Neal and Wheatley (1998), for 1994 through 1998 are from CDA/Wiesenberger, and for 1999 and 2000 from the discounts on stock funds reported in the *Wall Street Journal* in the turn-of-the-year issues.

Figure 3 shows the relationship between the dividend premium and the closed-end fund discount. They are not perfectly synchronous, but they are visibly related. The correlation is 0.37 with a p-value of 0.02. This provides some initial support for the sentiment mechanisms.

To tie this back to our basic results, Table 9 uses the closed-end fund discount as an instrumental variable for the dividend premium. The table also uses lagged capital gains and future relative returns on payers and nonpayers as instruments. The logic for using future relative returns is that they are arguably a purer (though perhaps noisier) measure of sentiment for dividends than the dividend premium. Recent capital gains on the market could capture either of the mechanisms outlined above – after a crash, unsophisticated investors may tend more toward the "bird in the hand" rationale, and also view general growth opportunities as bleak.[21]

Table 9 shows that the instrumental variables coefficients are, in general, about as statistically and economically significant as the basic OLS coefficients. For the specification that uses future returns as an instrument, this merely puts the earlier predictability results in units of the dividend premium. For the other specifications, the results have a more novel value. At a minimum, they confirm that that the specific component of the dividend premium associated with these variables helps to explain rates of initiation and omission, thus casting doubt on generic "omitted third factor" alternative explanations. To the extent that the instruments pick up investor sentiment, the results provide affirmative support for a sentiment interpretation.

VI. Conclusion

We develop a catering theory of dividends that focuses on the market efficiency assumption of the M&M dividend irrelevance proof. It adds to the collection of theories of

[21] We thank Lubos Pastor for suggesting that we use past capital gains in this manner.

dividends that relax other specific assumptions of the proof. The essence of catering is that managers give investors what they want. In the setting of dividends, catering implies that managers will tend to initiate dividends when investors put a relatively high stock price on dividend payers, and tend to omit dividends when investors prefer nonpayers. A simple model formalizes the key tradeoffs between maximizing fundamental value and catering, and offers testable time-series and cross-sectional predictions.

Our empirical work focuses on the central time-series prediction of the model: a positive relationship between the rates of dividend initiation and omission and the difference between the prevailing stock prices of payers and nonpayers. We test this relationship using four measures of investor demand for payers. The aggregate initiation rate is significantly positively related to all four (however, in one case this does not amount to more than a common trend). One proxy for investor demand for payers, the difference between the average market-to-book ratios of payers and nonpayers – the "dividend premium" – explains a statistically impressive three-fifths of the annual variation in the rate of initiation. In addition, the rate of omission is significantly negatively related to two of the four proxies for investor demand. After reviewing other possibilities, we conclude that catering is the most natural explanation.

We then inquire about the source of time-varying demand for dividends. We do not find strong evidence for a traditional dividend clientele. Instead, investor sentiment appears to affect the demand for dividends. This is suggested in the connection between the closed-end fund discount and the dividend premium, and in instrumental variables estimates of the effect of the dividend premium on dividend payment. In Baker and Wurgler (2002b), we review academic histories of the capital markets and historical financial news articles to construct a detailed timeline of how investor attitudes toward dividends have changed over time.

Appendix

This appendix describes the simulations that generate the bias-adjusted coefficients and p-values reported in Table 6. As discussed by Stambaugh (1999), a small-sample bias arises when the explanatory variable is persistent and there is a contemporaneous correlation between innovations in the explanatory variable and stock returns. For example, in the following system

$$R_t = a + bX_{t-1} + u_t \tag{A1}$$

$$X_t = c + dX_{t-1} + v_t\,, \tag{A2}$$

the bias is equal to

$$E[\hat{b} - b] = \tfrac{s_{uv}}{s_v^2} E[\hat{d} - d]\,, \tag{A3}$$

where the hats represent OLS estimates. Kendall (1954) shows the OLS estimate of d has a negative bias. The bias for OLS b is therefore of the opposite sign to the sign of the covariance between innovations in dividend policy and returns.

The sign of this covariance is not obvious a priori (unlike when the predictor is a scaled-price variable). To address the potential for bias and conduct inference, we use a bootstrap estimation technique. The approach is identical to Baker and Stein (2002) and is similar to that used in Vuolteenaho (2001), Kothari and Shanken (1997), Stambaugh (1999), and Ang and Bekaert (2001). For each regression in Table 6, we perform two sets of simulations.

The first set generates a bias-adjusted point estimate. We simulate (A1) and (A2) recursively starting with X_0, using the OLS coefficient estimates, and drawing with replacement from the empirical distribution of the errors u and v. We throw out the first 100 draws (to draw from the unconditional distribution of X), then draw an additional N observations, where N is the size of the original sample. (For the cumulative three-year regressions, the number of additional draws is one third the size of the original sample, since it contains overlapping returns.) With each simulated sample, we re-estimate (A1). This gives us a set of coefficients b^*. The bias-adjusted coefficient BA reported in Table 6 subtracts the bootstrap bias estimate (the mean of b^* minus the OLS *b*) from the OLS *b*.

In the second set of simulations, we redo everything as above under the null hypothesis of no predictability – that is, imposing *b* equals zero. This gives us a second set of coefficients b^{**}. With these in hand, we can determine the probability of observing an estimate as large as the OLS *b* by chance, given the true $b = 0$. These are the p-values in Table 6.

References

Allen, Franklin, Antonio E. Bernardo, and Ivo Welch, 2000, A theory of dividends based on tax clienteles, *Journal of Finance* 55, 2499-2536.

Allen, Franklin, and Roni Michaely, 2002, Payout policy, University of Pennsylvania working paper.

Alti, Ashiq, Hwang, Lee-Seok, and Mark A. Trombley, 2002, Arbitrage risk and the book-to-market mispricing, *Journal of Financial Economics* (forthcoming).

Ang, Andrew and Geert Bekaert, 2001, Stock return predictability: Is it there?, NBER working paper #8207.

Asquith, Paul, and David W. Mullins, Jr., 1983, The impact of initiating dividend payments on shareholders' wealth, *Journal of Business* 56, 77-96.

Baker, Malcolm, and Serkan Savasoglu, 2002, Limited arbitrage in mergers and acquisitions, *Journal of Financial Economics* (forthcoming).

Baker, Malcolm, and Jeremy C. Stein, 2002, Market liquidity as a sentiment indicator, Harvard University working paper.

Baker, Malcolm, Stein, Jeremy C., and Jeffrey Wurgler, 2001, When does the market matter? Stock prices and the investment of equity-dependent firms, Harvard University working paper.

Baker, Malcolm and Jeffrey Wurgler, 2000, The equity share in new issues and aggregate stock returns, *Journal of Finance* 55, 2219-2257.

Baker, Malcolm and Jeffrey Wurgler, 2002a, Market timing and capital structure, *Journal of Finance* 55, 2219-2257.

Baker, Malcolm and Jeffrey Wurgler, 2002b, Why are dividends disappearing? An empirical analysis, Harvard University working paper.

Baker, Malcolm, Robin Greenwood, and Jeffrey Wurgler, 2002, The maturity of debt issues and predictable variation in bond returns, *Journal of Financial Economics* (forthcoming).

Barberis, Nicholas, and Andrei Shleifer, 2002, Style investing, *Journal of Financial Economics* (forthcoming).

Barberis, Nicholas, Andrei Shleifer, and Robert W. Vishny, 1998, A model of investor sentiment, *Journal of Financial Economics* 49, 307-343.

Barberis, Nicholas, Andrei Shleifer, and Jeffrey Wurgler, 2001, Comovement, University of Chicago working paper.

Benartzi, Shlomo, Roni Michaely, and Richard Thaler, 1997, Do changes in dividends signal the future or the past?, *Journal of Finance* 52, 1007-1034.

Black, Fischer, and Myron S. Scholes, 1974, The effects of dividend yield and dividend policy on common stock prices and returns, *Journal of Financial Economics* 1, 1-22.

Black, Fischer, 1976, The dividend puzzle, *Journal of Portfolio Management*, 5-8.

Blanchard, Olivier, Chanyong Rhee, and Lawrence Summers, 1990, The stock market, profit, and investment, *Quarterly Journal of Economics* 108, 115-136.

Boehme, Rodney D., and Sorin M. Sorescu, 2002, The long-run performance following dividend initiations and resumptions: Underreaction or product of chance?, *Journal of Finance* 57, 871-900.

Brav, Alon, and J. B. Heaton, 1998, Did ERISA's prudent man rule change the pricing of dividend omitting firms?, Duke University working paper.

Brown, Stephen, and Jerold Warner, 1980, Measuring security price performance, *Journal of Financial Economics* 8, 205-258.

Campbell, John Y., Martin Lettau, Burton G. Malkiel, and Yexiao Xu, 2001, Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk, *Journal of Finance* 56, 1-44.

Campbell, John Y., Andrew W. Lo, and A. Craig MacKinlay, 1997, *The Econometrics of Financial Markets*, (Princeton University Press, Princeton, NJ).

Chen, Joseph, Harrison Hong, and Jeremy C. Stein, 2002, Breadth of ownership and stock returns, *Journal of Financial Economics* (forthcoming).

Cooper, Michael J., Orlin Dimitrov, and P. Raghavendra Rau, 2001, A rose.com by any other name, *Journal of Finance* 56, 2371-2388.

Daniel, Kent, Hirshleifer, David, and Avanidhar Subrahmanyam, 1998, Investor psychology and security market under- and overreactions, *Journal of Finance* 53, 1839-85.

D'Avolio, Gene, 2002, The market for borrowing stock, *Journal of Financial Economics* (forthcoming).

DeAngelo, Harry, Linda DeAngelo, and Douglas J. Skinner, 1996, Dividend signaling and the disappearance of sustained earnings growth, *Journal of Finance* 40, 341-371.

DeLong, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert Waldmann, 1990, Noise trader risk in financial markets, *Journal of Political Economy* 98, 703-738.

Del Guercio, Diane, 1996, The distorting effect of the prudent-man laws on institutional equity investments, *Journal of Financial Economics* 40, 31-62.

Duffie, Darrell, Nicolae Garleanu, and Lasse Heje Pedersen, 2002, Securities lending, shorting, and pricing, *Journal of Financial Economics* (forthcoming).

Eades, Kenneth M., Patrick J. Hess, and E. Han Kim, 1994, Time-series variation in dividend pricing, *Journal of Finance* 49, 1617-1638.

Fama, Eugene F., and Harvey Babiak, 1968, Dividend policy: An empirical analysis, *Journal of the American Statistical Association* 53, 1132-1161.

Fama, Eugene F., and Kenneth R. French, 2001, Disappearing dividends: Changing firm characteristics or lower propensity to pay?, *Journal of Financial Economics* 60, 3-44.

Feenberg, Daniel, and Elizabeth Coutts, 1993, An introduction to the Taxsim model, *Journal of Policy Analysis and Management* 12, 189-194.

Geczy, Christopher, David K. Musto, and Adam Reed, 2002, Stocks are special too: An analysis of the equity lending market, *Journal of Financial Economics* (forthcoming).

Gordon, Myron J., 1959, Dividends, earnings, and stock prices, *Review of Economics and Statistics* 41, 99-105.

Graham, Benjamin, and David L. Dodd, 1951, *Security Analysis: Principles and Techniques* (McGraw-Hill, New York, NY).

Graham, John R., and Campbell R. Harvey, 2001, The theory and practice of corporate finance: Evidence from the field, *Journal of Financial Economics* 60, 187-244.

Greenwood, Robin, 2001, Large events and limited arbitrage: Evidence from a Japanese stock index redefinition, Harvard University working paper.

Greenwood, Robin, and Nathan Sosner, 2001, Where do betas come from?, Harvard University working paper.

Grullon, Gustavo, and Roni Michaely, 2002, Dividends, share repurchases, and the substitution hypothesis, *Journal of Finance* (forthcoming).

Hakansson, Nils H., 1982, To pay or not to pay dividends, *Journal of Finance* 37, 415-428.

Healy, Paul M., and Krishna G. Palepu, 1988, Earnings information conveyed by dividend initiations and omissions, *Journal of Financial Economics* 21, 149-176.

Hong, Harrison, and Jeremy C. Stein, 1999, A unified theory of underreaction, momentum trading and overreaction in asset markets, *Journal of Finance* 54, 2143-2184.

Hubbard, Jeff, and Roni Michaely, 1997, Do investors ignore dividend taxation? A reexamination of the Citizens Utilities case, *Journal of Financial and Quantitative Analysis* 32, 117-135.

Hyman, Leonard, 1988, *America's Electric Utilities: Past, Present, and Future* (Arlington, VA: Public Utility Reports).

Jensen, Michael C., 1986, Agency costs of free cash flow, corporate finance and takeovers, *American Economic Review* 76, 323-329.

Jenter, Dirk, 2001, "Managerial portfolio decisions and market timing," Harvard University working paper.

John, Kose, and Joseph Williams, 1985, Dividends, dilution, and taxes: A signaling equilibrium, *Journal of Finance* 40, 1053-1070.

Kahneman, Daniel, and Amos Tversky, 1979, Prospect theory: An analysis of decision under risk, *Econometrica* 47, 263-291.

Kendall, M. G., 1954, Note on bias in estimation of auto-correlation, *Biometrika* 41, 403-404.

Kothari, S. P., and Jay Shanken, 1997, Book-to-market, dividend yield, and expected market returns: A time series analysis, *Journal of Financial Economics* 44, 169-203.

Lamont, Owen A., and Charles M. Jones, 2002, Short sale constraints and stock returns, *Journal of Financial Economics* (forthcoming).

Lamont, Owen A., and Richard H. Thaler, 2001, Can the market add and subtract? Mispricing in tech-stock carve-outs, University of Chicago working paper.

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert Vishny, 2000, Agency problems and dividend policies around the world, *Journal of Finance* 55, 1-33.

Lee, Charles M., Andrei Shleifer, and Richard Thaler, 1991, Investor sentiment and the closed-end fund puzzle, *Journal of Finance* 46, 75-110.

Lintner, John, 1956, The distribution of incomes among corporations among dividends, retained earnings, and taxes, *American Economic Review* 46, 97-113.

Long, John B., 1978, The market valuation of cash dividends: A case to consider, *Journal of Financial Economics* 6, 235-264.

Malkiel, Burton G., 1999, *A Random Walk Down Wall Street*, (Norton, New York, NY).

Marsh, Terry A., and Robert C. Merton, 1987, Dividend behavior for the aggregate stock market, *Journal of Business* 60, 1-40.

Mendenhall, Richard R., 2002, Post-earnings announcement drift and arbitrage risk, *Journal of Business* (forthcoming).

Miller, Merton H., and Franco Modigliani, 1961, Dividend policy, growth and the valuation of shares, *Journal of Business* 34, 411-433.

Miller, Merton H., and Kevin Rock, 1985, Dividend policy under asymmetric information, *Journal of Finance* 40, 1031-1051.

Miller, Merton H., and Myron Scholes, 1978, Dividends and taxes, *Journal of Financial Economics* 6, 333-364.

Mitchell, Mark, and Todd C. Pulvino, 2001, Characteristics of risk and return in risk arbitrage, *Journal of Finance* 56, 2135-2176.

Mitchell, Mark, Todd C. Pulvino, and Erik Stafford, 2002, Limited arbitrage in equity markets, *Journal of Finance* 57, 551-584.

Morck, Randall, Robert Vishny, and Andrei Shleifer, 1990, The stock market and investment: Is the market a sideshow?, *Brookings Papers on Economic Activity* 2:1990, 157-215.

Mullainathan, Sendhil, 2002, Thinking through categories, MIT working paper.

Myers, Stewart, 1984, The capital structure puzzle, *Journal of Finance* 39, 575-592.

Myers, Stewart, and Nicholas Majluf, 1984, Corporate financing and investment decisions when firms have information that investors do not have, *Journal of Financial Economics* 13, 187-221.

Neal, Robert, and Simon M. Wheatley, 1998, Do measures of investor sentiment predict returns?, *Journal of Financial and Quantitative Analysis* 33, 523-547.

Newey, Whitney K, and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703-708.

Ofek, Eli, and Matthew Richardson, 2002a, DotCom mania: The rise and fall of internet stock prices, *Journal of Finance* (forthcoming).

Ofek, Eli, and Matthew Richardson, 2002b, The valuation and market rationality of internet stock prices, NYU working paper.

Peterson, Pamela, David Peterson, and James Ang, 1985, Direct evidence on the marginal rate of taxation on dividend income, *Journal of Financial Economics* 14, 267-82.

Polk, Christopher, and Paola Sapienza, 2001, The real effects of investor sentiment, Northwestern University working paper.

Pontiff, Jeffrey, 1996, Costly arbitrage: Evidence from closed-end funds, *Quarterly Journal of Economics* 111, 1135-1152.

Pontiff, Jeffrey, and Michael J. Schill, 2001, Long-run seasoned equity offering returns: Data snooping, model misspecification, or mispricing? A costly arbitrage approach, University of Washington working paper.

Poterba, James M., 1986, The market valuation of cash dividends: The Citizens Utilities case reconsidered, *Journal of Financial Economics* 15, 395-405.

Rau, P. Raghavendra, Ajay Patel, Igor Osobov, Ajay Korana, and Michael J. Cooper, 2001, The game of the name: Value changes accompanying dot.com additions and deletions, Purdue University working paper.

Rosch, Eleanor, 1978, Principles of categorization, in Eleanor Rosch and Barbara B. Lloyd, eds.: *Cognition and Categorization* (Lawrence Erlbaum Associates, Hillsdale, NJ).

Shefrin, Hersh M., and Meir Statman, 1984, Explaining investor preference for cash dividends, *Journal of Financial Economics* 13, 253-282.

Shiller, Robert J., 1984, Stock prices and social dynamics, *Brookings Papers on Economic Activity* 2, 457-498.

Shiller, Robert J., 1989, *Market Volatility*, (MIT Press, Cambridge, MA).

Shiller, Robert J., 2000, *Irrational Exuberance,* (Princeton University Press, Princeton, NJ).

Shleifer, Andrei, and Robert W. Vishny, 1992, Equilibrium short horizons of investors and firms, *American Economic Review Papers and Proceedings* 80, 148-153.

Shleifer, Andrei, and Robert W. Vishny, 1997, The limits of arbitrage, *Journal of Finance* 52, 35-55.

Shleifer, Andrei, and Robert W. Vishny, 2002, Stock market driven acquisitions, Harvard University working paper.

Stambaugh, Robert F., 1999, Predictive regressions, *Journal of Financial Economics* 54, 375-421.

Stein, Jeremy C., 1989, Efficient capital markets, inefficient firms: A model of myopic corporate behavior, *Quarterly Journal of Economics* 104, 655-669.

Stein, Jeremy C., 1996, Rational capital budgeting in an irrational world, *Journal of Business* 69, 429-455.

Thaler, Richard H., 1980, Toward a positive theory of consumer choice, *Journal of Economic Behavior and Organization* 1, 39-60.

Thaler, Richard H., 1983, Using mental accounting in a theory of consumer behavior, Cornell University working paper.

Thaler, Richard H., and Hersh M. Shefrin, 1981, An economic theory of self-control, *Journal of Political Economy* 89, 392-406.

Vuolteenaho, Tuomo, 2000, Understanding the aggregate book-to-market ratio and its implications to current equity-premium expectations, Harvard University working paper.

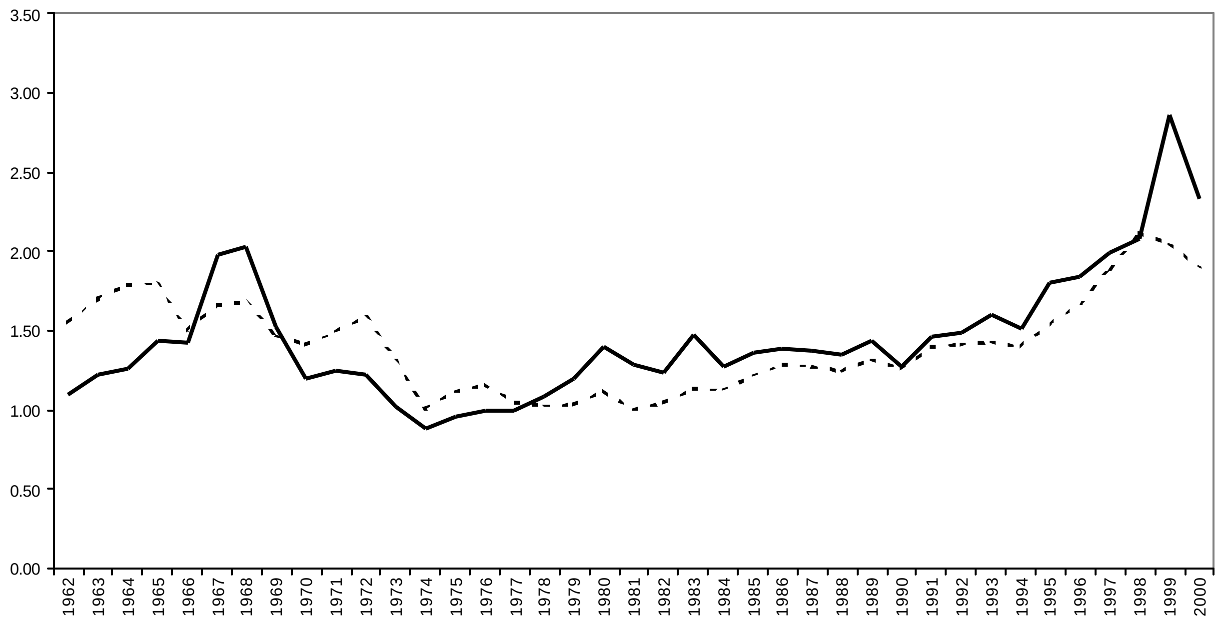
Watts, Ross, 1973, The information content of dividends, *Journal of Business* 46, 191-211.

Wurgler, Jeffrey, and Katia Zhuravskaya, 2002, Does arbitrage flatten demand curves for stocks?, *Journal of Business* 75, 583-608.

Zweig, Martin E., 1973, An investor expectations stock price predictive model using closed-end fund premiums, *Journal of Finance* 28, 67-87.

Figure 1. Valuation of dividend payers and nonpayers and the dividend premium, 1962-2000. The average market-to-book ratio for dividend payers and nonpayers and the dividend premium (the log difference in average market-to-book ratios). A firm is defined as a dividend payer at time *t* if it has positive dividends per share by the ex date (Item 26). The market-to-book ratio is the ratio of the market value of the firm to its book value. Market value is equal to market equity at calendar year end (Item 24 times Item 25) plus book debt (Item 6 minus book equity). Book equity is defined as stockholders' equity (generally Item 216, with exceptions as noted in the text) minus preferred stock (generally Item 10, with exceptions as noted in the text) plus deferred taxes and investment tax credits (Item 35) and post retirement assets (Item 330). The average market-to-book ratios are constructed by value-weighting (by book value) across dividend payers and nonpayers and are plotted in Panel A. Panel B plots the log difference between the market-to-book ratio of payers and nonpayers.

Panel A. Average market-to-book ratio of dividend payers (dash) and nonpayers (solid)



Panel B. The dividend premium %

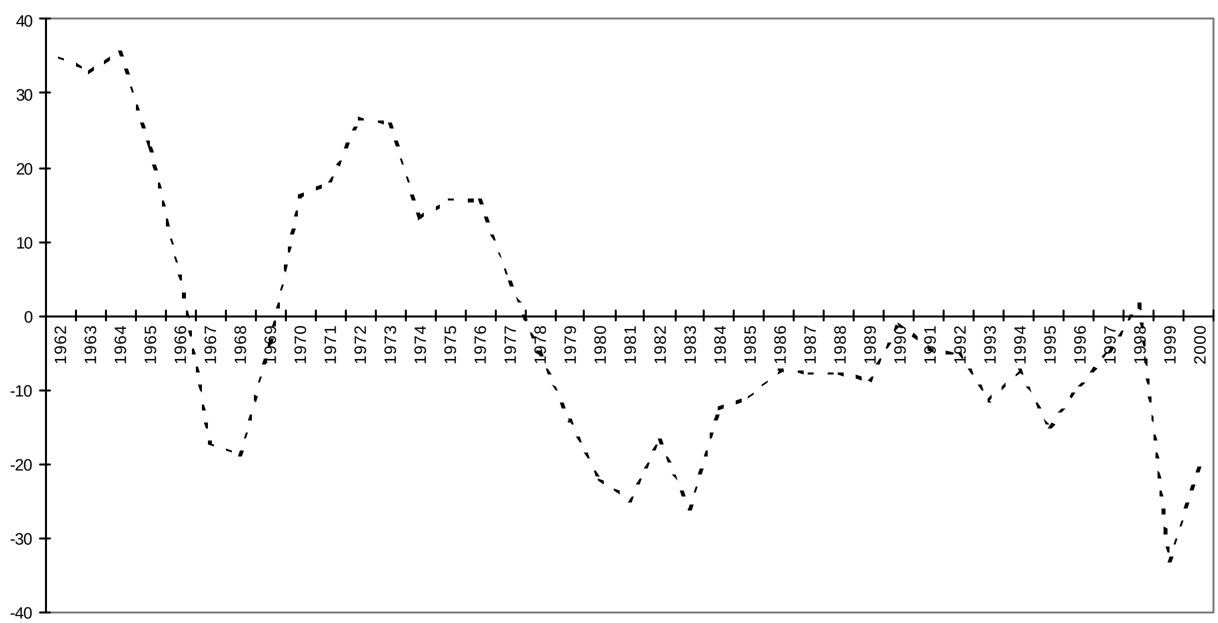
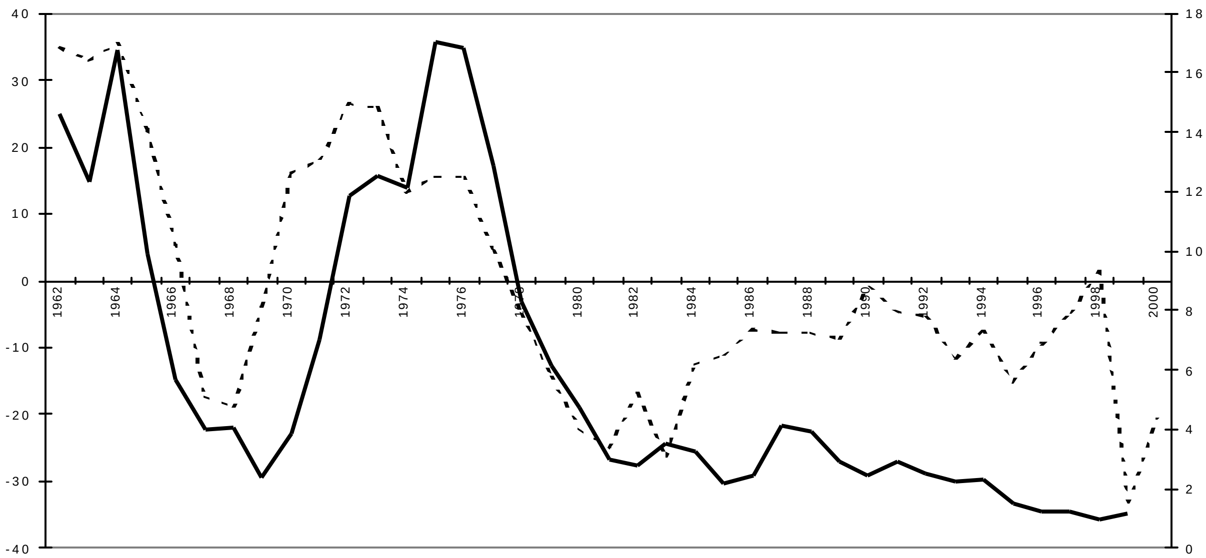


Figure 2. The dividend premium and the rate of dividend initiation, 1962-2000. The log difference in the market-to-book ratio of dividend payers and nonpayers (the dividend premium) and one-year-ahead dividend initiations. A firm is defined as a dividend payer at time *t* if it has positive dividends per share by the ex date (Item 26). Panel A plots the dividend premium against the raw initiation rate *Initiate* in $t+1$, defined as the number of new dividend payers at time $t+1$ among surviving nonpayers from t. Panel B plots the dividend premium against the estimated propensity to initiate $\widetilde{PTI}$ in $t+1$ (the rate of initiation controlling for prevailing firm characteristics in $t+1$). The propensity to initiate is estimated following Fama and French (2001). First, we estimate a set of annual Fama-MacBeth logit regressions of dividend payment on firm characteristics over the sample of surviving nonpayers, using as firm characteristics the NYSE market capitalization percentile *NYP*, the market-to-book ratio *M/B*, asset growth *dA/A*, and profitability *E/A*. Second, we define the propensity to initiate as the average annual prediction error (actual initiation rate minus predicted rate) of these regressions.

Panel A. *Initiate* (solid – right axis) and the dividend premium



Panel B. Propensity to initiate ($\widetilde{PTI}$) (solid – right axis) and the dividend premium

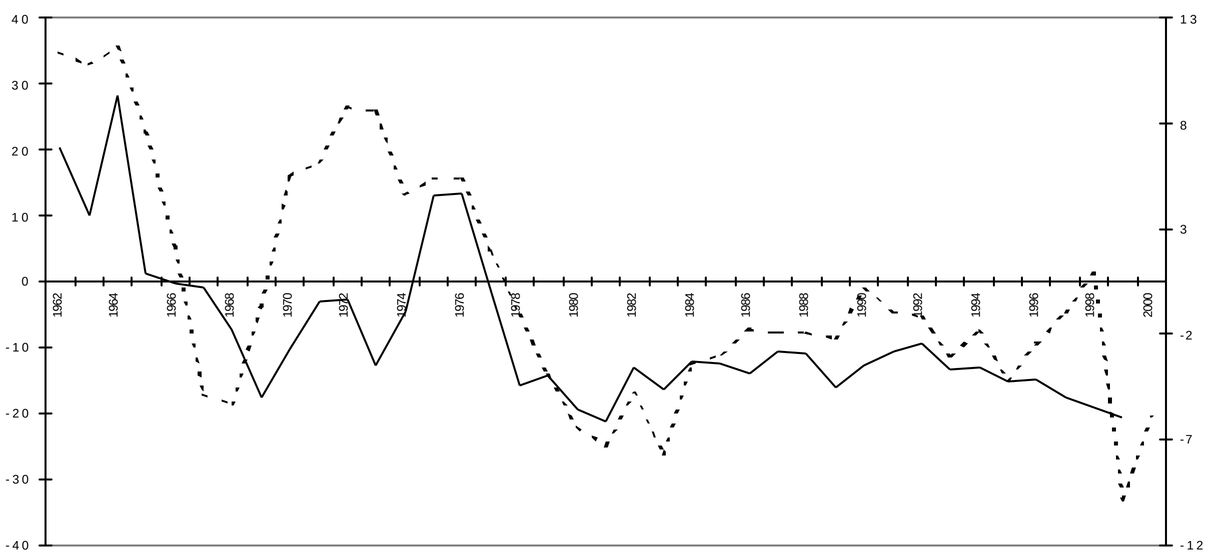


Figure 3. The dividend premium and the closed-end fund discount. The log difference in the market-to-book ratio of dividend payers and nonpayers (the dividend premium) and the closed-end fund discount. The value-weighted closed-end fund discount uses data on net asset values and market prices for general equity and convertible funds from Simon and Wheatley (1997) for 1962 to 1993, from CDA/Wiesenberger for 1994 to 1998, and from the *Wall Street Journal* for 1999 to 2000. The dividend premium (dash – left axis) is plotted against the contemporaneous closed end fund discount (solid – right axis).

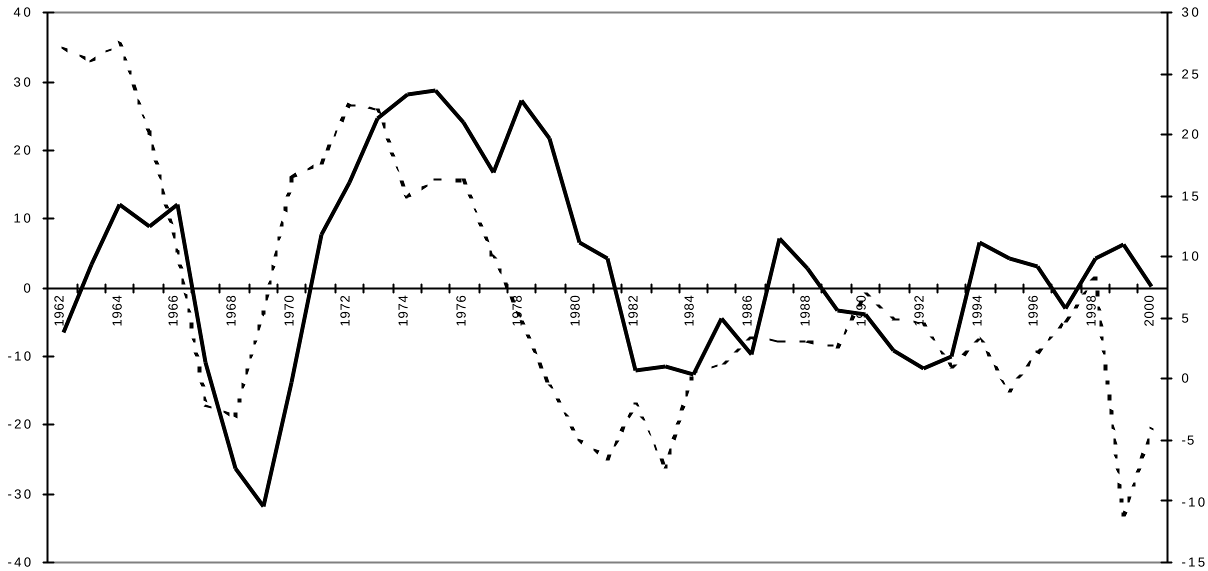


Table 1. Summary measures of dividend payment, 1962-2000. Dividend payers, nonpayers, and the rates at which subsamples pay dividends. A firm is defined as a dividend payer at time *t* if it has positive dividends per share by the ex date (Item 26). A firm is defined as a new dividend payer at time *t* if it has positive dividends per share by the ex date at time *t* and zero dividends per share by the ex date at time *t*-1. A firm is defined as an old payer at time *t* if it has positive dividends per share by the ex date at time *t* and positive dividends per share by the ex date at time *t*-1. A firm is defined as a new list payer if it has positive dividends per share by the ex date at time *t* and is not in the sample at time *t*-1. A firm is defined as a nonpayer at time *t* if it does not have positive dividends per share by the ex date. New nonpayers are firms who were payers at time *t*-1 but not at *t*. Old nonpayers a firms who were nonpayers in both *t*-1 and *t*. New list nonpayers are nonpayers at *t* who were not in the sample at *t*-1. The initiation rate *Initiate* expresses payers as a percentage of surviving nonpayers from *t*-1. The rate at which firms continue paying dividends *Continue* expresses payers as a percentage of surviving payers from *t*-1. The rate at which lists pay *Listpay* expresses payers as a percentage of new lists at *t*.

	Payers				*Nonpayers*				*Payment rates %*		
Year	**Total**	**New**	**Old**	**List**	**Total**	**New**	**Old**	**List**	**Initiate**	**Continue**	**Listpay**
1963	529	21	467	41	149	8	123	18	14.6	98.3	69.5
1964	585	17	519	49	154	6	121	27	12.3	98.9	64.5
1965	681	24	565	92	167	3	119	45	16.8	99.5	67.2
1966	821	16	659	146	238	5	145	88	9.9	99.2	62.4
1967	888	13	793	82	288	14	216	58	5.7	98.3	58.6
1968	954	11	849	94	361	19	263	79	4.0	97.8	54.3
1969	1,018	14	908	96	438	16	330	92	4.1	98.3	51.1
1970	1,048	10	946	92	554	54	406	94	2.4	94.6	49.5
1971	1,030	20	951	59	639	75	502	62	3.8	92.7	48.8
1972	1,281	43	953	285	862	52	568	242	7.0	94.8	54.1
1973	1,627	97	1,221	309	1,127	22	719	386	11.9	98.2	44.5
1974	1,719	130	1,535	54	1,044	44	908	92	12.5	97.2	37.0
1975	1,802	118	1,593	91	1,052	65	853	134	12.2	96.1	40.4
1976	1,878	167	1,670	41	941	58	813	70	17.0	96.6	36.9
1977	1,944	146	1,756	42	821	30	721	70	16.8	98.3	37.5
1978	1,956	96	1,747	113	856	53	651	152	12.9	97.1	42.6
1979	1,925	64	1,761	100	1,046	45	708	293	8.3	97.5	25.4
1980	1,854	58	1,735	61	1,137	68	882	187	6.2	96.2	24.6
1981	1,738	48	1,634	56	1,417	90	962	365	4.8	94.8	13.3
1982	1,631	37	1,545	49	1,621	78	1,210	333	3.0	95.2	12.8
1983	1,523	40	1,434	49	1,929	100	1,380	449	2.8	93.5	9.8
1984	1,450	59	1,346	45	2,111	50	1,605	456	3.5	96.4	9.0
1985	1,378	57	1,282	39	2,133	42	1,698	393	3.2	96.8	9.0
1986	1,270	39	1,176	55	2,373	73	1,744	556	2.2	94.2	9.0
1987	1,214	49	1,112	53	2,651	61	1,971	619	2.4	94.8	7.9
1988	1,185	92	1,057	36	2,563	50	2,123	390	4.2	95.5	8.5
1989	1,162	83	1,041	38	2,432	59	2,036	337	3.9	94.6	10.1
1990	1,148	61	1,053	34	2,403	49	2,011	343	2.9	95.6	9.0
1991	1,128	51	1,052	25	2,497	59	2,015	423	2.5	94.7	5.6
1992	1,140	62	1,036	42	2,674	56	2,085	533	2.9	94.9	7.3
1993	1,148	60	1,043	45	3,049	55	2,342	652	2.5	95.0	6.5
1994	1,163	61	1,059	43	3,286	55	2,634	597	2.3	95.1	6.7
1995	1,165	66	1,068	31	3,416	31	2,772	613	2.3	97.2	4.8
1996	1,153	44	1,061	48	3,774	40	2,924	810	1.5	96.4	5.6
1997	1,101	38	1,027	36	3,784	52	3,110	622	1.2	95.2	5.5
1998	1,042	37	978	27	3,501	35	2,997	469	1.2	96.5	5.4
1999	975	27	916	32	3,320	31	2,806	483	1.0	96.7	6.2
2000	871	30	824	17	3,042	50	2,587	405	1.1	94.3	4.0
Mean	1,247	55	1,119	73	1,693	45	1,336	312	6.1	96.2	27.0
SD	402	38	379	63	1,180	24	974	219	5.0	1.7	22.5

Table 2. The dividend premium, 1962-2000. The market valuations of dividend payers and nonpayers. A firm is defined as a dividend payer at time *t* if it has positive dividends per share by the ex date (Item 26). The market-to-book ratio is the ratio of the market value of the firm to its book value. Market value is equal to market equity at calendar year end (Item 24 times Item 25) plus book debt (Item 6 minus book equity). Book equity is defined as stockholders' equity (generally Item 216, with exceptions as noted in the text) minus preferred stock (generally Item 10, with exceptions as noted in the text) plus deferred taxes and investment tax credits (Item 35) and post retirement assets (Item 330). The market-to-book ratio is equal-weighed (EW) and value-weighted by book value (VW) across dividend payers and nonpayers. These ratios are calculated for the entire sample and for new lists. A firm is defined as a new list if it is not in the sample at time *t*-1. The dividend premium $P^{D\text{-}ND}$ is the difference between the logs of the dividend payers and nonpayers market-to-book ratios.

	Payers				*Nonpayers*				*Dividend Premium* ($P^{D\text{-}ND}$)			
	Total		List		Total		List		Total		List	
Year	**EW M/B**	**VW M/B**	**EW M/B**	**VW M/B**	**EW M/B**	**VW M/B**	**EW M/B**	**VW M/B**	**EW**	**VW**	**EW**	**VW**
1962	1.50	1.55	1.50	1.36	1.19	1.10	1.25	1.12	22.9	34.9	18.6	19.8
1963	1.58	1.70	1.71	1.49	1.30	1.23	1.88	1.71	19.4	32.9	-9.7	-13.8
1964	1.68	1.79	2.09	2.10	1.37	1.26	1.46	1.41	20.1	35.6	35.6	40.0
1965	1.76	1.80	1.60	1.47	1.61	1.43	1.74	1.52	8.8	22.6	-8.5	-3.1
1966	1.52	1.50	1.35	1.20	1.52	1.43	1.55	1.47	0.2	5.4	-14.3	-20.2
1967	1.87	1.66	2.34	1.83	2.36	1.98	3.42	2.65	-23.5	-17.2	-38.0	-36.8
1968	1.99	1.69	2.35	2.89	2.73	2.03	3.32	2.45	-31.7	-18.8	-34.4	16.8
1969	1.60	1.47	1.84	1.67	1.78	1.52	1.90	1.70	-10.4	-3.8	-3.4	-2.1
1970	1.43	1.41	1.51	1.67	1.38	1.20	1.77	1.64	3.1	16.0	-15.6	1.4
1971	1.64	1.50	2.14	2.01	1.48	1.25	2.23	1.90	10.3	18.2	-4.0	5.6
1972	1.62	1.59	1.70	1.74	1.48	1.22	1.84	1.47	9.4	26.6	-8.3	17.0
1973	1.19	1.32	1.27	1.27	1.16	1.02	1.46	1.27	3.2	25.9	-14.1	-0.7
1974	0.93	1.01	1.11	0.91	0.91	0.89	1.08	0.99	2.0	13.2	3.1	-7.6
1975	1.03	1.12	0.90	0.86	1.05	0.95	1.40	1.05	-2.5	15.6	-44.6	-19.9
1976	1.08	1.16	1.37	1.11	1.13	0.99	1.69	1.06	-4.2	15.6	-20.5	4.2
1977	1.06	1.05	1.24	1.23	1.18	1.00	1.32	1.09	-10.7	4.6	-6.3	12.0
1978	1.08	1.03	1.13	1.48	1.34	1.08	1.63	1.24	-22.1	-5.0	-36.5	17.6
1979	1.14	1.04	1.33	0.92	1.75	1.19	2.71	1.61	-43.2	-14.3	-71.6	-55.6
1980	1.25	1.12	1.87	1.20	2.33	1.40	3.86	1.69	-61.9	-22.1	-72.5	-34.2
1981	1.15	1.01	1.46	1.11	1.87	1.29	2.69	1.88	-48.2	-24.9	-61.2	-53.3
1982	1.23	1.05	1.37	1.32	2.03	1.24	3.14	2.05	-50.1	-16.9	-82.6	-44.1
1983	1.41	1.14	1.76	1.21	2.31	1.48	3.18	1.85	-49.3	-26.2	-59.1	-42.9
1984	1.31	1.13	1.72	1.47	1.79	1.28	2.29	1.41	-31.7	-12.5	-28.6	3.5
1985	1.43	1.21	1.64	0.91	2.00	1.36	3.07	1.82	-33.2	-11.0	-62.8	-68.6
1986	1.53	1.29	1.93	1.44	2.27	1.39	3.61	1.74	-39.7	-7.3	-63.0	-18.5
1987	1.47	1.28	1.85	1.53	2.03	1.38	2.83	1.55	-32.4	-7.8	-42.4	-1.4
1988	1.48	1.24	1.47	1.38	1.94	1.35	3.04	1.48	-27.2	-7.8	-72.8	-7.1
1989	1.54	1.32	1.51	1.25	1.97	1.44	3.08	1.61	-24.9	-8.7	-71.2	-25.3
1990	1.39	1.26	1.79	1.80	1.76	1.27	2.27	1.19	-23.5	-1.0	-23.5	41.4
1991	1.59	1.40	1.31	1.24	2.32	1.47	3.45	1.50	-37.8	-4.6	-96.8	-19.4
1992	1.63	1.41	2.03	1.34	2.23	1.49	2.82	1.72	-31.1	-5.3	-32.8	-25.1
1993	1.68	1.43	1.74	1.38	2.33	1.60	2.96	1.82	-33.1	-11.5	-53.2	-27.4
1994	1.55	1.40	1.48	1.47	2.04	1.51	2.59	1.82	-27.6	-7.5	-55.7	-21.7
1995	1.64	1.55	1.83	1.86	2.57	1.80	3.64	2.02	-44.7	-15.1	-68.6	-8.0
1996	1.69	1.67	2.05	1.88	2.41	1.84	3.03	2.09	-35.5	-9.4	-39.0	-10.7
1997	1.86	1.89	1.83	1.52	2.35	1.99	3.02	2.22	-22.9	-4.8	-50.1	-38.0
1998	1.79	2.12	1.98	2.21	2.22	2.09	3.57	2.17	-21.8	1.4	-59.0	1.9
1999	1.68	2.05	1.40	1.34	3.54	2.86	7.97	3.41	-74.9	-33.2	-173.6	-93.0
2000	1.65	1.90	2.18	1.48	2.26	2.33	3.03	1.69	-31.5	-20.6	-33.1	-13.3
Mean	1.48	1.42	1.66	1.48	1.88	1.45	2.64	1.69	-21.1	-0.7	-40.4	-13.6
SD	0.26	0.30	0.35	0.40	0.55	0.41	1.19	0.47	23.7	18.0	36.7	27.8

Table 3. The Citizens Utilities dividend premium and market reactions to dividend initiations, 1962-2000. The Citizens Utilities (CU) price ratio is the log of the ratio of the annual average cash dividend class share price to the annual average stock dividend class share price. The 1962 through 1972 data are from Long (1978) and the 1973 through 1989 data are from CRSP. A firm is defined as a new dividend payer at time *t* if it has positive dividends per share by the ex date (Item 26) at time *t* and zero dividends per share by the ex date at time *t*-1. We take the first dividend declaration date (DCLRDT) from CRSP in the twelve month period prior to the fiscal year ending in *t*. We calculate the sum of the differences between the firm return (RET) and the CRSP value-weighted market return (VWRETD) for a three-day window [-1, +1] around the declaration date. The announcement effect *A* scales this return by the standard deviation of the excess returns between 120 calendar days and five trading days before the declaration date. The test statistic from Campbell, Lo, and Mackinlay (1997, equation 4.4.24) is shown in braces and tests the null hypothesis of zero average price reaction in year *t*.

	CU Dividend Premium	*Initiation Announcement Effect*			
Year	P^{CU}	**N**	**Excess Return**	*A*	**[t-stat]**
1962	0.96	1	5.40	1.75	[1.73]
1963	0.98	17	1.94	0.47	[1.92]
1964	1.00	21	1.70	0.41	[1.85]
1965	1.00	21	1.43	0.40	[1.81]
1966	1.00	10	-0.84	-0.23	[-0.73]
1967	0.95	10	0.18	0.06	[0.19]
1968	0.97	7	2.20	0.54	[1.40]
1969	0.97	10	1.82	0.37	[1.16]
1970	1.00	8	5.46	0.85	[2.37]
1971	0.96	19	2.08	0.37	[1.60]
1972	0.93	39	2.17	0.51	[3.14]
1973	0.96	112	3.45	0.70	[7.33]
1974	0.99	94	5.92	0.87	[8.34]
1975	0.96	128	5.21	0.77	[8.59]
1976	0.93	128	4.97	1.05	[11.75]
1977	0.91	114	4.28	1.12	[11.82]
1978	0.90	68	4.02	0.79	[6.43]
1979	0.89	43	3.62	0.70	[4.53]
1980	0.87	35	3.50	0.58	[3.38]
1981	0.92	33	3.57	0.89	[5.08]
1982	0.93	22	3.93	0.62	[2.89]
1983	0.81	25	3.49	0.85	[4.24]
1984	0.89	47	2.13	0.42	[2.85]
1985	0.93	34	1.25	0.35	[2.04]
1986	1.00	31	3.17	0.51	[2.80]
1987	0.92	50	1.38	0.16	[1.15]
1988	0.86	65	2.11	0.48	[3.86]
1989	0.84	50	3.68	0.78	[5.50]
1990	.	46	5.85	0.74	[4.96]
1991	.	31	5.20	0.63	[3.50]
1992	.	46	2.53	0.50	[3.39]
1993	.	42	0.55	0.06	[0.41]
1994	.	51	0.94	0.21	[1.50]
1995	.	44	1.81	0.39	[2.58]
1996	.	18	6.24	0.86	[3.61]
1997	.	20	2.35	0.52	[2.33]
1998	.	19	0.93	0.20	[0.87]
1999	.	17	2.38	0.28	[1.15]
2000	.	10	4.78	0.81	[2.54]
Mean	0.94	41	2.99	0.57	[3.48]
SD	0.05	33	1.75	0.35	[2.87]

Table 4. Statistics for demand for dividend measures, 1962-2000. The first column shows the autocorrelation coefficient, the second column shows a Dickey-Fuller test, and the remaining columns show the correlations among the variables. The dividend premium $P^{D\text{-}ND}$ is the difference between the logs of the EW and VW market-to-book ratios for dividend payers and nonpayers. The Citizens Utilities dividend premium P^{CU} is the log of the ratio of the annual average cash dividend class share price to the annual average stock dividend class share price. The initiation announcement effect A is the average standardized excess return in a three-day window [-1, +1] around the first declaration dates by new dividend payers. Future relative returns $r_{Dt+1} - r_{NDt+1}$ is the difference in real returns for value-weighted indexes of dividend payers and nonpayers in year t+1. Future relative returns $R_{Dt+3} - R_{NDt+3}$ is the cumulative difference in future returns from year t+1 through t+3. P-values are in brackets.

			Dividend premium				*Future returns*	
	r	**Unit Root**	**VW**	**EW**	P_t^{CU}	A_t	$r_{Dt+1} - r_{NDt+1}$	$R_{Dt+3} - R_{NDt+3}$
VW P_t^{D-ND}	0.82	-1.98	1.00					
	[0.00]	[0.29]						
EW P_t^{D-ND}	0.82	-1.58	0.95	1.00				
	[0.00]	[0.49]	[0.00]					
P_t^{CU}	0.61	-2.00	0.60	0.63	1.00			
	[0.00]	[0.28]	[0.00]	[0.00]				
A_t	0.40	-5.18	0.25	0.18	-0.20	1.00		
	[0.02]	[0.00]	[0.13]	[0.27]	[0.31]			
$r_{Dt+1} - r_{NDt+1}$	0.10	-5.31	-0.21	-0.24	-0.28	0.16	1.00	
	[0.54]	[0.00]	[0.20]	[0.15]	[0.14]	[0.35]		
$R_{Dt+3} - R_{NDt+3}$	0.70	-2.52	-0.54	-0.47	-0.28	-0.19	0.63	1.00
	[0.00]	[0.11]	[0.00]	[0.00]	[0.15]	[0.27]	[0.00]	

Table 5. Dividend payment and demand for dividends: Basic relationships, 1962-2000. Regressions of initiation and omission rates on measures of the dividend premium. For example, the initiation rate is modeled in Panel A as:

$$Initiate_t = a + bP_{t-1}^{D-ND} + cA_{t-1} + dP_{t-1}^{CU} + u_t$$

The initiation rate *Initiate* expresses payers as a percentage of surviving nonpayers from *t*-1. The continuation rate *Continue* expresses payers as a percentage of surviving payers from *t*-1. The rate at which listing firms pay *Listpay* expresses payers as a percentage of new lists at *t*. The dividend premium P^{D-ND} is the difference between the logs of the EW and VW market-to-book ratios for dividend payers and nonpayers. The announcement effects *A* are the average standardized excess returns in a three-day window [-1, +1] around the declaration dates of new dividend payers. The Citizens Utilities dividend premium P^{CU} is the log of the ratio of the annual average cash dividend class share price to the annual average stock dividend class share price. The independent variables are standardized to have unit variance. T-statistics use standard errors that are robust to heteroskedasticity and serial correlation up to four lags.

	(1)	**(2)**	**(3)**	**(4)**	**(5)**
			Panel A: $Initiate_t$		
VW P_{t-1}^{D-ND}	3.90				3.80
	[6.56]				[10.74]
EW P_{t-1}^{D-ND}		3.63			
		[5.10]			
P_{t-1}^{CU}			1.70		-0.52
			[2.21]		[-0.82]
A_{t-1}				2.15	1.06
				[2.51]	[1.52]
N	38	38	28	38	28
R^2	0.60	0.52	0.11	0.18	0.70
			Panel B: $Continue_t$		
VW P_{t-1}^{D-ND}	0.85				1.00
	[2.83]				[2.59]
EW P_{t-1}^{D-ND}		0.93			
		[2.96]			
P_{t-1}^{CU}			0.44		-0.25
			[1.02]		[-0.61]
A_{t-1}				0.03	-0.24
				[0.09]	[-0.87]
N	38	38	28	38	28
R^2	0.26	0.30	0.06	0.00	0.25
			Panel C: $Listpay_t$		
VW P_{t-1}^{D-ND}	16.08				10.11
	[6.29]				[2.12]
EW P_{t-1}^{D-ND}		18.15			
		[7.12]			
P_{t-1}^{CU}			14.74		8.16
			[4.68]		[1.64]
A_{t-1}				2.98	-0.28
				[0.58]	[-0.11]
N	38	38	28	38	28
R^2	0.51	0.65	0.47	0.02	0.63

Table 6. Dividend payment and demand for dividends: Predicting returns, 1962-2000. Univariate regressions of future excess returns of dividend payers over nonpayers on the initiation rate, the continuation rate, and the rate at which listing firms pay. The dependent variable in Panel A is the difference in real returns between dividend payers r_D and nonpayers r_{ND}. The dependent variable in Panel B is real return of dividend payers r_D. The dependent variable in Panel C is the real return of nonpayers r_{ND}. R_{t+k} denotes cumulative returns from t+1 through t+k. The initiation rate *Initiate* expresses new payers as a percentage of surviving nonpayers from t-1. The continuation rate *Continue* expresses continuing payers as a percentage of surviving payers from t-1. The rate at which listing firms pay *Listpay* expresses new Compustat lists who are payers as a percentage of new Compustat lists. In the *Listpay* specification, a year trend is included in the regression. The independent variables are standardized to have unit variance. We report OLS coefficients and bias-adjusted (BA) coefficients. Bootstrap p-values represent a two-tailed test of the null hypothesis of no predictability.

		Initiate$_t$				*Continue$_t$*				*Listpay$_t$ (detrended)*			
	N	**OLS**	**BA**	**[p-val]**	**R^2**	**OLS**	**BA**	**[p-val]**	**R^2**	**OLS**	**BA**	**[p-val]**	**R^2**
Panel A: Relative returns													
$r_{Dt+1} - r_{NDt+1}$	37	-7.68	-6.54	[0.15]	0.10	-7.68	-7.97	[0.06]	0.10	-6.13	-6.87	[0.16]	0.07
$r_{Dt+2} - r_{NDt+2}$	36	-13.27	-12.63	[0.01]	0.31	-7.90	-8.20	[0.07]	0.11	-9.47	-9.49	[0.03]	0.15
$r_{Dt+3} - r_{NDt+3}$	35	-8.81	-8.79	[0.06]	0.14	-5.90	-6.13	[0.17]	0.07	-7.08	-7.49	[0.09]	0.08
$R_{Dt+3} - R_{NDt+3}$	35	-30.54	-28.23	[0.06]	0.47	-21.62	-23.63	[0.13]	0.25	-24.88	-23.91	[0.10]	0.28
Panel B: Payer returns													
r_{Dt+1}	37	-4.06	-4.39	[0.29]	0.06	-2.14	-2.41	[0.46]	0.02	-3.11	-2.15	[0.36]	0.03
r_{Dt+2}	36	-0.95	-1.67	[0.79]	0.00	0.70	0.54	[0.85]	0.00	-3.00	-2.51	[0.31]	0.03
r_{Dt+3}	35	-1.87	-2.28	[0.60]	0.01	1.12	0.92	[0.73]	0.00	-2.88	-2.85	[0.34]	0.02
R_{Dt+3}	35	-8.08	-10.71	[0.39]	0.10	-0.16	-0.23	[0.99]	0.00	-8.19	-4.83	[0.37]	0.09
Panel C: Nonpayer returns													
r_{NDt+1}	37	3.62	2.26	[0.64]	0.01	5.54	5.76	[0.38]	0.03	3.01	4.93	[0.62]	0.01
r_{NDt+2}	36	12.32	11.02	[0.07]	0.13	8.60	8.73	[0.16]	0.07	6.47	6.83	[0.26]	0.03
r_{NDt+3}	35	6.94	6.54	[0.31]	0.04	7.02	7.27	[0.28]	0.04	4.20	4.80	[0.47]	0.01
R_{NDt+3}	35	22.46	17.45	[0.23]	0.18	21.47	24.81	[0.21]	0.17	16.70	19.88	[0.36]	0.09

Table 7. Dividend payment and the dividend premium: Other controls, 1962-2000. Regressions of dividend payment rates on measures of the dividend premium, growth opportunities, the personal tax advantage of dividends versus capital gains, and a time trend. For example, the initiation rate is modeled in Panel A as:

$$Initiate_t = a + bP_{t-1}^{D-ND} + c\frac{M}{B}_{t-1} + d\frac{D}{P}_{t-1} + eTax_{t-1} + fYear_{t-1} + u_t$$

The initiation rate *Initiate* expresses payers as a percentage of surviving nonpayers from t-1. The continuation rate *Continue* expresses payers as a percentage of surviving payers from t-1. The rate at which listing firms pay *Listpay* expresses payers as a percentage of new lists at t. The dividend premium P^{D-ND} is the difference between the logs of the VW market-to-book ratios for dividend payers and nonpayers. The VW market-to-book ratio *M/B* is averaged across nonpayers in Panels A, payers in Panel B, and new lists in Panel C. The VW dividend yield *D/P* is from CRSP. *Tax* is the ratio of after-tax income from a dollar in dividends to a dollar in long-term capital gains. All independent variables but *Year* are standardized to unit variance. T-statistics use standard errors that are robust to heteroskedasticity and serial correlation up to four lags.

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
			Panel A: *Initiate*$_t$			
VW P_{t-1}^{D-ND}	2.83	2.71	2.87	4.19	3.66	3.90
	[5.39]	[5.42]	[5.42]	[6.53]	[7.65]	[4.56]
VW *Nonpayer* M/B_{t-1}	-1.92	-1.34	-1.32			
	[-2.43]	[-2.54]	[-2.32]			
VW D/P_{t-1}				1.63	0.95	0.96
				[3.05]	[1.90]	[2.13]
Tax_{t-1}		1.48	1.72		1.37	1.74
		[3.22]	[2.28]		[2.64]	[2.02]
$Year_{t-1}$			0.03			0.05
			[0.40]			[0.52]
N	38	38	38	38	38	38
R^2	0.70	0.77	0.77	0.70	0.75	0.76
			Panel B: *Continue*$_t$			
VW P_{t-1}^{D-ND}	0.79	0.57	0.45	0.83	0.56	0.40
	[2.64]	[2.30]	[1.75]	[2.64]	[2.19]	[1.56]
VW *Payer* M/B_{t-1}	0.30	0.50	0.48			
	[1.05]	[2.02]	[2.28]			
VW D/P_{t-1}				-0.16	-0.50	-0.50
				[-0.82]	[-1.85]	[-2.43]
Tax_{t-1}		0.60	0.39		0.68	0.43
		[2.37]	[1.62]		[2.47]	[2.05]
$Year_{t-1}$			-0.03			-0.04
			[-0.78]			[-0.95]
N	38	38	38	38	38	38
R^2	0.29	0.39	0.40	0.27	0.38	0.39
			Panel C: *Listpay*$_t$			
VW P_{t-1}^{D-ND}	16.88	13.86	5.85	16.35	10.92	2.60
	[7.75]	[7.31]	[3.55]	[5.67]	[5.39]	[2.62]
VW *List* M/B_{t-1}	2.89	5.36	3.93			
	[0.76]	[2.85]	[3.13]			
VW D/P_{t-1}				1.54	-5.34	-5.54
				[0.47]	[-2.51]	[-5.57]
Tax_{t-1}		12.29	-0.19		13.99	1.05
		[6.50]	[-0.11]		[6.36]	[0.97]
$Year_{t-1}$			-1.67			-1.79
			[-7.37]			[-15.74]
N	38	38	38	38	38	38
R^2	0.53	0.79	0.95	0.52	0.78	0.96

Table 8. Dividend payment and the dividend premium: Firm characteristics controls, 1963-2000. Two-stage regressions of dividend payment on firm characteristics and the dividend premium. The first stage performs a set of Fama-MacBeth logit regressions of dividend payment on firm characteristics for each of three samples: surviving payers, surviving nonpayers, and new lists.

$$\Pr\left(Payer_{it}=1\right)=\text{logit}\left(a+bNYP_{it}+c\frac{M}{B}_{it}+d\frac{dA}{A}_{it}+e\frac{E}{A}_{it}\right)+u_{it}$$

The second stage regresses the average annual prediction errors (actual policy minus predicted policy) from the logit regressions on the dividend premium. For example, for the sample of surviving nonpayers we estimate the residual rate of initiation, or the "propensity to initiate dividends" *PTI*, and model it as:

$$P\widetilde{T}I_t = f + gP_{t-1}^{D-ND} + v_t, \text{ where } P\widetilde{T}I_t \equiv \tfrac{1}{N}\sum_i u_{it}\ .$$

The first two rows examine the propensity to initiate *PTI* and so restrict the sample to surviving nonpayers. The next two rows examine the propensity to continue *PTC* and so restrict the sample to surviving payers. The last two rows examine the propensity to list as a payer *PTL* and so restrict the sample to new lists. The firm characteristics are the NYSE percentile *NYP*, the market-to-book ratio *M/B* (which is excluded in some specifications), asset growth *dA/A*, and profitability *E/A*. The NYSE percentile is the percentage of firms listed on the NYSE that are equal to or smaller in terms of market capitalization (PRC*SHROUT). The market-to-book ratio is the ratio of the market value of the firm to its book value. Market value is equal to market equity at calendar year end (Item 24 times Item 25) plus book debt (Item 6 minus book equity). Book equity is defined as stockholders' equity (generally Item 216) minus preferred stock (generally Item 10) plus deferred taxes and investment tax credits (Item 35) and post retirement assets (Item 330). Asset growth is the change in assets (Item 6) over assets. Profitability is earnings before extraordinary items (Item 18) plus interest expense (Item 15) plus income statement deferred taxes (50) over assets. The dividend premium $P^{D\text{-}ND}$ is the difference between the logs of the VW market-to-book ratios for dividend payers and nonpayers. T-statistics in the second stage regression use standard errors that are robust to heteroskedasticity and serial correlation up to four lags.

	NYP_t		M/B_t		dA/A_t		E/A_t		VW P_{t-1}^{D-ND}	
	b	**[t-stat]**	**c**	**[t-stat]**	**d**	**[t-stat]**	**e**	**[t-stat]**	**g**	**[t-stat]**
PTI	1.54	[10.65]	-0.85	[-6.62]	0.21	[1.13]	9.60	[10.00]	3.35	[6.66]
PTI	0.90	[6.13]			-0.23	[-1.48]	6.47	[12.42]	3.76	[6.71]
PTC	4.57	[10.09]	0.33	[1.31]	1.50	[5.15]	15.06	[5.87]	0.34	[1.73]
PTC	4.61	[10.63]			1.37	[4.96]	14.20	[6.01]	0.32	[1.56]
PTL	4.56	[40.95]	-0.78	[-15.86]	-0.84	[-6.44]	10.76	[11.67]	11.20	[5.51]
PTL	3.88	[37.16]			-1.19	[-8.64]	7.80	[13.06]	12.78	[6.99]

Table 9. Dividend payment and the dividend premium: Instrumental variables, 1962-2000. Instrumental variables estimates of the effect of the dividend premium on dividend payment rates. For example, the initiation rate is modeled in Panel A as:

$$Initiate_t = a + b\hat{P}_{t-1}^{D-ND} + u_t$$

$$P_{t-1}^{D-ND} = c + dCEF_{t-1} + e_1 r_{CGt-1} + e_2 r_{CGt-2} + e_3 r_{CGt-3} + f\left(R_{Dt+3} - R_{NDt+3}\right) + v_{t-1}$$

The initiation rate *Initiate* expresses payers as a percentage of surviving nonpayers from t-1. The continuation rate *Continue* expresses payers as a percentage of surviving payers from t-1. The rate at which listing firms pay *Listpay* expresses payers as a percentage of new lists at t. The dividend premium VW P^{D-ND} is the difference between the logs of the value-weighted average market-to-book ratios of dividend payers and nonpayers. We instrument for the dividend premium with the value-weighted closed-end fund discount, past nominal capital gains on the value-weighted CRSP index, and future cumulative relative returns of payers over nonpayers. The value-weighted closed-end fund discount uses data on net asset values and market prices for general equity and convertible funds from Simon and Wheatley (1997) for 1962 to 1993, from CDA/Wiesenberger for 1994 to 1998, and from the *Wall Street Journal* for 1999 to 2000. r_{CGt-k} denotes capital gains in year t-k. R_{t+k} denotes cumulative future returns from t+1 through $t+k$. The OLS t-statistics use standard errors that are robust to heteroskedasticity and serial correlation up to four lags. The 2SLS t-statistics use standard errors that are robust to heteroskedasticity.

	OLS	*2SLS*			
		CEF Discount	**Past Capital Gains**	**Future Returns**	**All**
		Panel A: *Initiate*$_t$			
VW P_{t-1}^{D-ND}	3.90	8.31	6.25	5.51	5.42
	[6.56]	[3.86]	[3.85]	[4.71]	[6.60]
N	38	38	38	36	36
R^2	0.60	-	-	-	-
		Panel B: *Continue*$_t$			
VW P_{t-1}^{D-ND}	0.85	1.63	0.65	1.60	1.15
	[2.83]	[1.82]	[1.25]	[2.71]	[3.52]
N	38	38	38	36	36
R^2	0.26	-	-	-	-
		Panel C: *Listpay*$_t$			
VW P_{t-1}^{D-ND}	16.08	8.72	26.99	11.47	17.69
	[6.29]	[1.10]	[4.28]	[1.97]	[5.34]
N	38	38	38	36	36
R^2	0.51	-	-	-	-