The Effect of Dividend Initiations on Stock Returns: A Propensity Score Matching Approach

Yanli Wang

April 3, 2005

Abstract

This paper measures the effect of dividend initiation announcements on firms' stock returns using a propensity score matching approach. Unlike the traditional event study methodology, propensity score matching can reduce the bias in the estimation of dividend initiation effects by controlling for the existence of confounding factors. Consistent with previous studies, the results show that dividend initiations have significantly positive effects on stock returns. More interestingly, the reaction of stock returns to dividend initiation announcements exhibits apparent heterogeneity. The overall empirical evidence presented in the paper supports a tradeoff model based on the benefits and costs of dividend payments.

1 Introduction

Dividend policy is one of the most controversial issues in corporate finance. The conflict between the apparent tax disadvantage and the ubiquity of dividend payouts has puzzled economists for many years. This paper uses a propensity score matching approach to provide empirical support for a tradeoff model based on the benefits and costs of dividends.

The costs of dividend payouts are relatively easy to measure and calculate for managers. The initiation of dividends usually entails a double tax, a shortfall in capital resources, high flotation costs of new security issues, a future payout commitment, management fees, legal fees, and other associated costs. Given adequate accounting and market information, managers can reasonably measure the costs firms will incur in the initiation of dividends.

Compared with the costs of dividends, their benefits are more difficult to identify. Economists have for a long time tried to identify the role of dividends, but have not reached a consensus. The dividend irrelevance proposition of Miller and Modigliani (1961) provides a benchmark for research on dividend policy. They demonstrate that in a perfect market, dividend policy does not affect a firm's value and is therefore irrelevant. Since then, several theories have been developed. Bhattacharya (1979), Asquith and Mullins (1986), Ofer and Thakor (1987), John and Williams (1985), and Miller and Rock (1985) propose a signaling hypothesis. They argue that dividends represent favorable signals about the future prospects of firms. This hypothesis on the information content of dividends is also addressed in Miller and Modigliani (1961). Meanwhile, Rozeff (1982), Easterbrook (1984) and Jensen (1986) provide agency cost explanations for dividends arguing that dividends reduce agency costs. Although dividend payments cause a shortfall in capital resources as well as high flotation costs of raising external funds, they serve as a monitoring mechanism. With its commitment to a regular dividend payout, a firm has to make frequent visits to the capital market to raise new funds. This implies that it comes under the increased scrutiny and supervision of the SEC, banks, other stakeholders, brokers and other concerned parties; this reduces agency cost. In the absence of the dividend commitment, boards of directors would have to hire outsiders, e.g., auditors, to monitor the management, which tends to be expensive. Furthermore, dividend payments may reduce free cash flow and consequently prevent consumption of perquisites by managers and overinvestment. Thus, according to the above literature, the possible benefits of dividend payments are the signaling of a firm's future prospects and the reduction of agency costs. In addition, different firms have different clientele (Miller and Scholes (1978)). For institutional investors and individual investors in low tax bracket, they also have tax advantage. While neither of these benefits of dividends is easily measurable, the market reaction to dividend payments provides an indirect measure of these benefits. Thus, one straightforward way of measuring the net benefit of dividends is by observing market reactions to dividend announcements, especially to dividend initiation announcements in the sense that measurement on dividend initiation effects would minimize the bias of expectation since the firm has no history of dividend payments.

That there are positive stock price reactions to dividend initiations is widely accepted in the empirical literature in finance. Asquith and Mullins (1983) investigated 168 firms that initiated dividends during the period 1963 to 1980 and reported a 3.7 percent cumulative excess return over a 2-day announcement period. The results also show that the positive excess returns are positively related to the size of the initial payment. Healy and Palepu (1988) confirm the significantly positive impact of dividend initiations on stock returns and also find that firms that initiate dividends have significant increases in their earnings for at least the year prior to, the year of, and the year following dividend initiation. Mickaely, Thaler and Womack (1995) test both short-run and long-run effects of dividend initiations on stock returns and report a 3.4 percent excess return over a three-day horizon and a much larger excess return in post-dividend initiation years.

Since the previous empirical evidence [1] shows that dividend initiation announcements have positive effects on stock returns, the question arises as to why I want to measure the impact of such announcements? The reasons are two-folded. First, most of the extant papers focus on average reactions, and ignore the apparent heterogeneity in reactions of stocks to dividend initiations. For instance, Asquith and Mullins (1983) report that 30 percent of firms experienced negative stock reactions to dividend initiations. They claim that such a negative effect may be the result of the costs associated with the initial payments, such as an increase in tax burdens, transaction, and administration costs of the dividend program. Jin (2000) focuses on

[1] Benesh et al. (1984), Venkatesh (1989) and Rimbey and Officer (1992) also report similar results.

this issue and explores dividend initiations during the period 1973-1993. He shows that dividend initiating firms fall into two categories or types of firms, with initiation being a value-increasing event for one category and a value-decreasing event for the other. According to these results, the positive average stock reaction to dividend initiation is not applicable to a substantial number of firms and that firms should not simply assume that dividend initiations would increase their market value.

Second, all previous empirical work employs the event study methodology to test for the dividend initiation effect; however, this methodology is plagued by a growing number of anomalies. Event studies have been widely adopted in financial economics to investigate market responses to new information. Economists use a variety of market models to estimate normal returns and calculate the abnormal returns which are the actual post-announcement return of the security over the event window minus the normal return of the security. The unbiased abnormal return should be the return difference caused solely by the dividend event announcement. Any return changes caused by other characteristics, such as firm size and book-to-market ratio, should be excluded. Unfortunately, single-factor models (the CAPM or other market models) fail to do this. In fact, the evidence against CAPM has been growing since the late 1970s. Fama and French (1992) provide a detailed discussion of the anomaly literature and conclude that the beta in the CAPM model does not predict cross-section stock returns and that firm size, leverage, book-to-market ratio and the earning-to-price ratio all help to explain returns better. In addition, the expected return and beta measures are both very sensitive to the choice of market portfolio. Moreover, the constant term in market models would be biased too since the macroeconomic environment during the estimation period could be very different from the post-event period. Thus, the estimates of normal returns in event studies tend to be biased.

Apparently, Fama and French (1992) suggest a multifactor model. Following the analysis on the five risk factors, Fama and French (1993) propose a three factor model in which the beta, size and book and market ratio is controlled. They rank the size and split all the stocks into two groups: small and big; then they also break the stocks into three book-to-market equity groups: low (bottom 30%), medium(middle 40%) and high (top 30%). After all these ranking, six portfolios are constructed. The excess return of the ith security over the risk free rate is a linear function of the market risk premium, the difference between the returns on a portfolio of small stocks and large stocks, and the difference between the returns on a portfolio of high and low book-to-market ratio. The three-factor model does a good job in describing the average returns, while it fails to capture the continuation of short term returns. Fama and French (1996) make a thorough discussion on the multifactor model and suggest that future work should look for a richer model with additional risk factors.

The three-factor model is similar to that of traditional matching, which directly constructs the counterfactual ('normal return') by assuming that the untreated outcome is unrelated to the treatment (announcement) status conditional on some set of observed variables X. Thus, with a similar characteristics set X, the *ex post* return of a non-dividend announcement firm can be an unbiased estimate of the normal return of firms with a dividend announcement. Unfortunately, it is difficult to match on multiple dimensions, especially when X consists of continuous variables. This is the so-called “curse of dimensionality.”

Rosenbaum and Rubin (1983) propose a propensity-score matching approach to overcome this dimensionality problem. They match the treatment and comparison units with a function of all characteristic variables, which is nothing but a balancing score given that the distribution of characteristic variables is the same for the treatment and comparison groups. Briefly, this method summarizes all the characteristics into a single index making multi-dimensional matching possible. In addition, the flexibility of the logit (or probit) specification allows for a reduction in the bias generated by unobservable confounding factors. Compared to the event study approach, the propensity score matching approach has fewer restrictions and assumptions. There is no assumption that return distributions are normal, no assumptions on functional form, and most importantly, there are more factors combined to explain the cross-sectional effect on returns. Theoretically, the propensity score matching is a more unbiased and efficient way of estimating the “counterfactual.” A couple of papers have applied the propensity-score matching method to non-experimental causal studies, such as Heckman et al. (1997, 1998), Dehejia and Wahba (1999, 2002), and Smith and Todd (2003) among others. Financial economists have also started to use this approach to measure the impact of financial events. For example, Cheng (2003) and Li and Zhao (2003) have applied propensity score matching to investigate the long-run stock performance after seasoned equity openings (SEOs).

In this paper, I use the propensity-score matching approach to test the effect of dividend initiation announcements on stock returns. The set of characteristic covariates X consists of the following six variables: market beta, firm size, leverage ratio, market-to-book ratio, and earnings-to-price ratio and industry dummies. As I know, this is the first time that PSM is applied into this area and so many risk factors are combined into the return models. All of these variables appear to influence both dividend initiation decisions and stock returns. The nearest neighborhood matching method is adopted to estimate the effect on stock returns (the average effect of treatment on the treated, i.e., ATT) over three different event windows 0, (-1, 0) and (-1, 1). Consistent with the previous empirical evidence, the results show a significantly positive effect of dividend initiation announcements on stock returns during the period 1988-2003. Moreover, the estimated ATT exhibit heterogeneity across different years. All these empirical results supports the tradeoff model based on the benefits and costs of dividend payments. Finally, although PSM also has its own weakness, it is superior to the market and three-factor models since it can reduce the bias generated by the existence of confounding factors.

The rest of this paper is organized as follows. Section 2 provides a theoretical review of the policy evaluation methods, such as the event study and propensity score matching techniques. Section 3 gives a detailed description of methodology and estimation procedures, and Section 4 lists data employed in the paper and sample statistics. The results are presented and discussed in Section 5. Section 6 concludes the paper.

2 Policy Evaluation Problems

To evaluate the impact of a policy intervention, of a program or an event (i.e., treatment), the natural way is to measure the difference between the outcomes observed for the units with the treatment and the outcomes had there been no treatment. Assume that R_1 is the outcome under the treatment and that R_0 is the outcome without treatment. Accordingly, the impact or the causal effect for a specific unit i can be written as

$$\Delta_i = R_{i1} - R_{i0} .$$

The Average Effect of Treatment on the Treated (or ATT) is

$$E(\Delta_i \mid D_i = 1) = E(R_{i1} \mid D_i = 1) - E(R_{i0} \mid D_i = 1) ,$$

where $D = \{0,1\}$ is the indicator of the state. If unit i is assigned to treatment, $D_i = 1$; otherwise $D_i = 0$. $E(R_{i0} \mid D_i = 1)$ can be interpreted as the estimated outcome that would have been observed for unit i had it not been treated. In the program evaluation literature, it is called the "counterfactual." It is obvious that only R_{i1} or R_{i0} can be observed for each unit i, which makes $R_{i0} \mid D_i = 1$ unobservable. This is the key issue in any evaluation problem. To solve this problem, one needs to find an unbiased estimator of $R_{i0} \mid D_i = 1$. Sections 2.1, 2.2 & 2.3 present three different methods to estimate this counterfactual.

2.1 The Market Model

In financial economics, event studies are widely used in investigating stock market responses to public announcements of new information. In this method, the estimator of $R_{i0} \mid D_i = 1$ is termed "normal return," and is estimated by the single-factor model (e.g., the market model). From the perspective of program evaluation, event studies construct individual-specific impacts and then aggregate them to obtain the average treatment effect. The market model is probably the most frequently used approach, and is expressed as follows:

$$R_{it} = \alpha_i + \beta_i R_{mt} + \varepsilon_{it},$$

where α_i is a constant term for the ith stock , β_i is the market beta of the ith stock, R_{mt} is the market return, and ε_{it} is an error term. The parameters of the models are estimated by time-series data from the estimation period that precedes each individual announcement. The estimated parameters are then matched with actual returns over the event period. Thus, the abnormal returns (AR) in the equation can be calculated from actual returns during the event period and the estimated coefficients from the estimation period:

$$AR_{it} = R_{it} - \hat{\alpha}_i - \hat{\beta}_i R_{mt},$$

which is the estimated impact of the treatment.

In the market model, the risk of the firm is related to a single factor beta. Evidence against this single-factor model has been growing since the late 1970s. Factors such as size, earning-to-price ratio, and market-to-book ratio have been discovered to have an effect on the average asset return, indicating important deviations from the CAPM and related single factor models. Accordingly, firm-size-adjusted models and multifactor models have emerged to extend the theory to account for multiple sources of risk.

2.2 The Three-Factor Model

Fama and French (1992) make a summary on the anomalies against single factor model and suggest a multifactor model. Following the analysis on the five risk factors, Fama and French (1993) propose a three factor model in which the beta, size and book and market ratio is controlled. They rank the size and split all the stocks into two groups: small and big; then they also break the stocks into three book-to-market equity groups: low (bottom 30%), medium (middle 40%) and high (top 30%). After all these ranking, six portfolios are constructed. The excess return of the ith security over the risk free rate is a linear function of the market risk premium , the difference between the returns on a portfolio of small stocks and large stocks , $E(SMB)$, and the difference between the returns on a portfolio of high and low book-to-market ratio, $E(HML)$:

$$E(R_i) - R_f = b_i[E(R_m) - R_f] + s_i E(SMB) + h_i E(HML) .$$

The three-factor model does a better job in describing the average returns than the CAPM or market models; however, it fails to capture all the anomalies. Fama and French (1996) make a thorough discussion on the multifactor model and suggest that future work should look for a richer model with additional risk factors.

2.3 Matching Method

Instead of constructing individual-specific impacts, the matching method constructs conditional means on the set of covariates X. Based on this method, the evaluation problem is solved by assuming that the selection is not related to the untreated outcome conditional on some set of observed variables X. The basic assumption underlying matching is the conditional independence assumption (CIA). It can be denoted as

$$(R_0, R_1) \perp D \mid X \ ,$$

which says the treatment status is random conditional on some set of observed variables X. The CIA will be satisfied if X includes all of the variables that affect both participation and outcome. The intuition is that units with the same X have the same probability of treatment and that the treatment status is randomly assigned. If CIA is satisfied, it can solve the sample selection problem and identify the counterfactual as follows:

$$\begin{aligned} \Delta &= E_X\left[E(R_1 \mid D=1, X) - E(R_0 \mid D=1, X)\right] \\ &= E_X\left[E(R_1 \mid D=1, X) - E(R_0 \mid D=0, X)\right]. \end{aligned}$$

The other condition that has to be satisfied is the common support constraint,

$$0 < \Pr(D=1 \mid X) < 1 .$$

This condition ensures that each treatment unit can be paired with a corresponding control unit. Compared with a linear regression, the matching method needs no linear functional form. As long as the CIA holds, the matching estimator is consistent. Unlike a linear regression, matching needs no assumptions on the distribution of the error term.

Basically both the market model and the three-factor model are doing matching, too. The market model is matching on Beta, while three-factor model are also matching on size and book-to-market ratio. To capture all anomalies, more factors should be added into the model; however, incorporating more factors would give rise to the so-called "curse of dimensionality", which is the key problem in matching methods. Obviously, this method can only be applied to the case with a discrete set X. If there are many characteristics variables (n) and each variable has many discrete values (m), then one needs to match n to m cells. In reality, most variables are continuous, so it is impossible to implement one by one matching. This is where single-index matching proves useful.

2.3 Propensity score matching (PSM)

To match multiple characteristics variables simultaneously, Rosenbaum and Rubin (1983) suggest a propensity-score matching (PSM) method in which treatment units and control units are matched by a propensity score *p(X)* which satisfies:

$$X \perp D \mid p(X).$$

This propensity score is actually *a function of the observed covariates X such that the conditional distribution of X given p(X) is the same for treated (D=1) and control (D=0) units.* Specifically, it is the probability of assignment of the treatment to the units with some specific characteristics X,

$$p(X)=pr(D=1|X).$$

Based on the two identification conditions in the traditional matching method, Rosenbaum and Rubin (1983) prove that if treatment status is random conditional on X and if treatment and control units have a common support, that is,

$$(R_0, R_1) \perp D \mid X \text{, and } 0 < \Pr(D=1 \mid X) < 1 \text{ for all X,}$$

then the randomization also holds conditionally on the propensity score p(X); and for a specific $p(X)$, both treatment and control units can be found, that is,

$$(R_0, R_1) \perp D \mid p(X) \quad \text{and} \quad 0 < \Pr(D = 1 \mid p(X)) < 1 \quad \text{for all X.}$$

Based on these results, the ATT can be rewritten as

$$\begin{aligned}\Delta &= E_{p(X)}\left[E(R_1 \mid D = 1, p(X)) - E(R_0 \mid D = 1, p(X))\right] \\ &= E_{p(X)}\left[E(R_1 \mid D = 1, p(X)) - E(R_0 \mid D = 0, p(X))\right]\end{aligned}$$

Thus, PSM reduces multiple dimensions to one dimension.

Rosenbaum and Rubin (1983) develop this method to deal with the selection bias problem in observational studies. In the evaluation problem, data often come from non-randomized observational studies, not from randomized experimental studies. The method can reduce the bias in the estimation of treatment effects with observational data sets by controlling for the existence of the confounding factors.

The weakness of PSM is that it is not robust for the case that there are outcome unobservables dependent on D even conditioning on X. This happens when the agents (making treatment decision) have private information unobservable to the investors or analysts and lead to a bias on ATT, which is addressed in Heckman and Navarro-Lozano (2003); however, PSM does reduce, even though not eliminates, the bias generated by unobservable confounding factors based on the idea that the bias is reduced when the comparison of outcomes is performed using treated and control subjects who are as similar as possible. Thus, in this sense, PSM is still superior to the market and multifactor models since it combines more risk factors.

3 Methodology

In section 2, the rationale of propensity-score matching was briefly introduced. In order to apply this method to the estimation of the dividend initiation announcement effect, the basic model first needs to be constructed.

3.1 Model construction

The event or treatment considered in this paper is dividend initiation. Accordingly, the treatment group is composed of firms initiating dividends (D firms) while the comparison group consists of all non-dividend-initiating firms (ND firms). The outcome of interest is stock returns at the dividend declaration date. Specifically, the final variable of interest is ATT denoted by Δ :

$$\begin{aligned}\Delta &= E(R_1 \mid D=1) - E(R_0 \mid D=1) \\ &= E_{p(X)}\left[E(R_1 \mid D=1, p(X)) - E(R_0 \mid D=0, p(X))\right].\end{aligned}$$

To derive this, one first has to select the reasonable covariates X that determine both treatment decisions and stock returns, and specify the logistic function to satisfy the balance property. With the common support constraint, a matching with the control group can be made in terms of the estimated propensity score. Finally, one chooses an appropriate matching method to estimate the counterfactual and calculate ATT.

3.2 Variable Selection

The CIA, i.e., $(R_0, R_1) \perp D \mid X$, can be satisfied only if X includes all variables that affect both the dividend initiation decisions and the outcomes, stock returns. Variables that affect dividend initiation decisions but not stock returns would not lessen the selection bias but would worsen the support problem. In a non-experimental evaluation problem there is no systematic mechanism to select variables. However, the selection is not arbitrary. Based on finance theories and previous empirical evidence, there are a number of variables commonly accepted as playing roles in dividend payout decisions.

Earnings constrain firms' dividend paying capacity. Lintner (1956) shows that firms gradually adjust dividends in response to changes in earnings. Fama and Babiak (1968) extend Lintner's model by incorporating a lagged earnings variable and show that both earnings factors have a significant effect on dividend changes. These two papers show that firms tend to smooth their dividend payments from year to year and link them to long-run earnings. Since the present paper is only concerned with dividend initiations, only short-run earnings are considered here. Fama and French (2001) show that firm size, profitability and investment opportunities affect the decision to pay dividends. In their paper, profitability is measured as common stock earnings over book equity since common stock earnings are more relevant to dividend decisions. Thus, earnings are a fundamental determinant of dividend policy. In the present paper, the earning-price ratio (*EPR*) is used as the proxy for earnings. Size is widely accepted as a state variable of a firm and as a general risk factor; it is a good proxy for the degree of publicly available information about a firm. Fama and French (2001) find that larger firms are more willing to pay dividends. In the present paper, size is measured by the natural logarithm of the market value of equity.

The firm's growth opportunity is another important determinant of its dividend initiation decisions. In the finance literature, Tobin's q is widely used as a proxy for growth opportunities. The basic idea behind Tobin’s q is that the firm should acquire more assets when its Tobin's q exceeds one, and should not acquire new assets unless it can create at least as much market value as the cost of reproducing them. Thus, a high Tobin’s q indicates good growth opportunities, while a low Tobin’s q implies poor or unrecognized growth opportunities. Intuitively, firms with high growth opportunities will tend to retain earnings to finance new projects, and will have little incentive to pay dividends. The most common estimate of Tobin’s q is the firm’s market-to-book ratio, which is defined as the market value of total assets divided by the book value of total assets. Fama and French (2001) use the market-to-book ratio to measure investment opportunities and show that it is one of the three factors that affect dividend payout decisions.

In agency cost theory in finance, both financial leverage and dividends are treated as mechanisms that can reduce agency costs. Easterbrook (1984) argues that dividends compel firms to frequently resort to the capital market to raise new funds, bringing the operations of firms and managers under increased scrutiny, thereby reducing managerial opportunism. Financial leverage also has a bearing on dividend policy. As the debt-to-equity ratio increases, bondholders limit dividends to prevent a wealth transfer from bondholders to shareholders; therefore, firms with higher leverage will be less likely to pay dividends.

The relationship between beta and dividend payments was first explored in Rozeff (1982). Rozeff argues that the tradeoff between the decline in agency costs engendered by increased dividends and the increase in the transaction costs of external financing produce a unique optimum for dividend payments. Since these transaction costs are directly related to the risk factors associated with operating and financial leverage, high beta is a reflection of high operating and financial leverage and therefore, beta has a negative correlation with dividend payouts. The empirical results in Rozeff's paper confirm this hypothesis. Dyl and Hoffmeister (1986) demonstrate that a firm's dividend policy will affect the duration of common shares and, consequently, will affect the riskiness of the firm's stock. In view of this, they argue that firms may pay dividends in order to select a 'preferred habitat' with respect to the riskiness of the common shares. Since beta is a good proxy for the systematic risk of the firm, it can be an important factor influencing the dividend initiation decision.

I have argued so far that the following factors, market beta, firm size, financial leverage, Tobin's q and firm earnings, are expected to affect the dividend initiation decision. In the literature, additional variables such as free cash flow, dispersion of ownership, institutional holdings, and earnings volatility have also been deemed to affect dividend policy. To be incorporated into the model, these factors must also affect stock returns, however; otherwise, the inclusion of these variables in the model would only cause worse support problems, as I discussed before. Thus, whether these factors have influence on stock returns should be further explored.

The CAPM asserts that security betas are sufficient to explain the cross-section of expected security returns. However, since the late 1970s, it has been shown that additional factors have explanatory power in determining expected returns. Fama and French (1992a) propose the following factors: firm size, leverage, earning-price ratio and book-to-market ratio[2]. Fama and French also explore the joint roles of market beta, size, leverage, book-to-market value of equity and earning-price ratio in explaining the cross-section of average stock returns. They find that used alone or in combination with other variables, beta provides little information about average returns. While size, earning-to-price ratio, leverage, and book-to-market ratio all have strong explanatory power when used alone. Fama and French (1993) identify three common risk factors as determinants of stock returns: an overall market factor, firm size, and book-to-market ratio. The overall market factor (market beta) appears to capture the time-varying aspect of stock returns, while firm size and book-to-market ratio explain the cross-sectional variations of stock returns.

Based on the above discussion, in this paper five covariates are included in the characteristic set X: market beta, firm size, leverage ratio, market-to-book ratio[3], and earning-to-price ratio. All of them affect both the dividend initiation decision and the cross-section of stock returns. Despite the negligible roles of leverage and the earning-to-price ratio when they are used in combination, I include them as the covariates for stock returns, for two reasons. First, the linear specification of the two factors is probably wrong. Second, the high order or interaction specification may reveal unobserved factors and reduce the bias generated by them, which is a key

[2] Not also that, Banz (1981) documents the firm-size effect; Bhandari (1988) finds the positive relationship between leverage and average return; Basu (1983) shows empirically that earning-price ratios help explain the cross section of average returns; Stattman (1980) finds that the average return is positively correlated with the market-to-book ratio.

[3] Instead of book-to-market ratio, I use market-to-book ratio as one of the characteristic variables because it is a more reasonable explanatory factor for dividend initiation decisions.

feature of the propensity score matching approach. In addition, based on the literature, dividend payout shows different patterns across different industries and returns also could be influenced by industrial factors, so I add an industrial variable into X. Thus, I have totally six characteristic variables.

Fama and French (2001) have shown that firms have become less likely over time to pay dividends. Baker and Wurgler (2002) develop a theory arguing that managers cater to investors' preferences on dividend policy and that the decision to pay dividends is driven by investor demand. Both papers suggest that dividend policy varies with investors' preferences, which are time-varying. To capture this time-varying feature of dividend-paying behavior, we do the regression year by year to avoid the side effects of variations in the economic environment.

3.3 Estimating the propensity scores: p(X)

After selecting the characteristic variables, we proceed to estimate the propensity scores. Rosenbaum and Rubin (1983) define a propensity score as a function of the observed covariates X such that the conditional distribution of X given $p(X)=Pr(D=1|X)$ is the same for the treated (D=1) and control (D=0) units, that is, $X \perp D \mid p(X)$ or $E(D \mid X, p(X)) = E(D \mid p(X))$. The intuition behind this balancing property is the following: after conditioning on $Pr(D = 1|X)$, additional conditioning on X should not provide new information about dividend initiation. Otherwise, if D is still dependent on X, a misspecification of the covariates X in the model used to estimate P(X) is implied. The most frequently used model is logit or probit. In this paper, the logit model is employed to estimate the propensity of dividend initiation, which is expressed as:

$$p(X_i) = \Pr(D_i = 1 | X_i) = \frac{\exp(\beta g(X_i))}{1 + \exp(\beta g(X_i))}$$

where g(X) is not necessarily linear with the covariates X.

Since the balancing condition is crucial for the conditional independence assumption (CIA) in propensity score matching, a balance test is necessary for the model specification. The idea is to test whether or not there are differences in X between the D = 1 and D = 0 groups after conditioning on P(X). Rosenbaum and Rubin (1985) suggest a measure based on standardized differences between the treatment and matched comparison group samples in terms of the means of each variable in X, the squares of each variable in X, and the first-order interaction terms between each pair of variables in X. Some papers use a variant of this measure to carry out the test. An alternative approach (Dehejia and Wahba (2002)) divides the observations into strata based on the estimated propensity scores. These strata are chosen so that there is no statistically significant difference in the means of the estimated propensity scores between the treatment and comparison group observations within each stratum. Then, within each stratum, t-tests are used to test for mean differences in each X variable between the treatment and comparison group observations. When significant differences are found for particular variables, higher order and interaction terms in those variables are added to the logistic model and the testing procedure is repeated until such differences no longer emerge.

Based on Dehejia and Wahba (2002), a variant of the algorithm for estimating the propensity score is implemented in this paper as follows:

1. Start estimating the logit model with a parsimonious specification, which means that only linear terms are included.
2. Split the sample into k equal intervals of the estimated propensity score; I start with k=5.
3. Within each interval, test that the average propensity scores of treated and control units do not differ.
4. If the test fails in one interval, split the interval into halves and test again. Continue until, in all intervals, the average propensity score of treated and control units do not differ.
5. Within each interval, test that the means of each characteristic do not differ between treated and control units. This is the necessary condition for the balance property.
6. If the means of one or more characteristics differ, that is, the balancing properties are not satisfied, higher-order terms and first-order interaction terms should be added.

This is a recursive procedure and should not stop until the balancing property is satisfied. It is worth mentioning that steps 2 to 5 are restricted to the common support, where the supports of the estimated propensity scores in the D and ND groups overlap. Imposing the common support condition in the estimation of the propensity scores may improve the quality of the matches used to estimate the ATT. The details are discussed in Section 3.4.

3.4 Solving the support problem

The second key condition that needs to be satisfied in propensity score matching is the common support constraint, $0 < \Pr(D = 1 | X) < 1$, for all X. By definition, the region of common support includes only those values of *P(X)* that have a positive density within both groups. To ensure that each D unit has a corresponding matching unit in the comparison group, the densities should be strictly positive and exceed zero by a threshold amount determined by a "trimming level" q. Heckman, Ichimura, and Todd (1997) exclude regions of the propensity score that have an estimated density below a cutoff value in either the treated or the untreated samples. In this paper, the simple algorithm suggested in Heckman, Ichimura, and Todd (1997) is adopted as follows:

Step 1: Find the overlap region over which both the D and ND propensity score densities are positive. Assume $\hat{S}_1$ and $\hat{S}_0$ are, respectively, the estimated smoothed support of the propensity scores of D firms and ND firms; thus, the estimated common support region is:

$$\hat{S}_{10} = \{P \in \hat{S}_1 \cap \hat{S}_0 : \hat{f}(P | D = 1) > 0 \ \ and \ \ \hat{f}(P | D = 0) > 0\}.$$

Step 2: find the region over which both the D and ND propensity score densities are strictly positive, above a trimming level q,

$$\hat{S}_q = \{P \in \hat{S}_{10} : \hat{f}(P | D = 1) > c_q \ \ \text{and} \ \ \hat{f}(P | D = 0) > c_q\},$$

where c_q is the density cut-off trimming level and determined by

$$\sup_{c_q} \frac{1}{2J} \sum_{i \in I_1} \{1(\hat{f}(P_i | D = 1) < c_q) + 1(\hat{f}(P_i | D = 0) < c_q)\} \leq q,$$

where I_1 is the set of observed P of the D firms that lie in $\hat{S}_{10}$ and *J* is the number of the observations in I_1. Notice that P_i is the estimated propensity score of the D firms;

thus, this algorithm deletes the D firms that have few counterparts in the ND group or have a very low occurrence rate in D group.

3.5 Matching on p(X)

After the propensity scores are estimated and the two identification conditions are satisfied, I construct the control group and estimate the ATT. Smith and Todd (2003) provide a comprehensive discussion of different ATT estimators and matching methods. In their paper, the cross-section estimator of ATT is used:

$$\Delta = \frac{1}{n_1} \sum_{i \in I_1 \cap S_{10}} \left[R_{i1} - E(R_{i0} \mid D_i = 1, p(X_i) \right],$$

where $E(R_{i0} \mid D_i = 1, p(X_i)) = \sum_{j \in I_0} W(i, j) R_{j0}$ and I_0 is the set of observed P of the ND firms that lie in $\hat{S}_{10}$. Different matching methods differ in how they construct the weights *W (i, j).* Three issues are involved in the matching. First, whether to match with replacement or not. With replacement, the score distance is minimized and the bias is reduced. Without replacement, the bias will increase while the variance decreases. Second, how many comparison units are there to match to each treated unit? With more units, the precision increases since more information is incorporated. With fewer units, the estimator is less biased. Finally, which matching method is appropriate? This choice depends on the data. In this paper, I adopt nearest neighbor matching, which is the most commonly used method in the literature. After the ATT is derived, a test is made to see if this effect is significant. A detailed discussion is given in Section 5.

4 Data and Sample Selection

The dividend initiation observations are drawn from the CRSP monthly stock files which covers all firms listed in the NYSE, AMEX and NASDAQ during the period 1988-2003. The corresponding *ex-ante* characteristic data are extracted from the merged CRSP/COMPUSTAT annual industrial files. To avoid a serious selection bias, several screening criteria are imposed on the sample selection. For each year t,

1. All the shares selected (in the treatment and comparison groups) are ordinary common shares with a share code of 10 or 11.
2. Owing to complications due to regulations, firms in the financial service sector (SIC: 6000-6999) and in the utility sector (SIC: 4000-4949) are excluded.
3. The distribution must be ordinary dividends, which are paid regularly in cash. Extra or special, year-end or final, interim and non-recurring dividends are excluded.
4. There are no other distribution announcements in a one-month window; otherwise the confounding effects would contaminate the results.
5. All the screened firms (dividend-initiating and non-dividend-initiating) must be listed in Compustat. For each screened firm, the Compustat must have the following characteristics data at the end of year t-1: total assets (data 6), shares outstanding (data 25), earning-per-share (data 58), liabilities (data 181), stock price (data 199), and book equity (data 216).

6. For each screened firm, the data on market beta must be available in the CRSP database by the end of year t-1.

Finally, these 7 criteria result in 744 dividend-initiating cases for firms listed before year t. The initiation rates for firms newly listed are also reported. The detailed sample information is listed in Table 1.

Table 1: Dividend Initiation Sample Description (1988-2003)

New lists are firms that get listed in the CRSP database between June of year t-1 and May of year t. Old lists are firms that are added to Compustat before fiscal year t and have corresponding *ex ante* characteristic data in Compustat by the end of fiscal year t-1. Old lists must be in the CRSP database.

	New Lists			*Old Lists*		
Year	*New lists*	*Initiation*	*Percent*	*Old lists*	*Initiation*	*Percent*
1988	468	42	8.97	2704	91	3.37
1989	345	32	9.28	2691	77	2.86
1990	361	27	7.48	2522	62	2.46
1991	291	10	3.44	2559	40	1.56
1992	571	36	6.30	2473	45	1.82
1993	550	37	6.73	2715	50	1.84
1994	836	46	5.50	2884	39	1.35
1995	555	28	5.05	3275	48	1.47
1996	867	41	4.73	3382	26	0.77
1997	818	34	4.16	3730	32	0.86
1998	666	25	3.75	3781	20	0.53
1999	400	14	3.50	3854	31	0.80
2000	700	16	2.29	3676	22	0.60
2001	416	15	3.61	3764	17	0.45
2002	176	27	15.34	3581	29	0.81
2003	277	129	46.57	3044	115	3.78
Total	8297	559	-	50635	744	-
Average	519	35	8.54	3164	47	1.47

Consistent with the results of Fama and French (2001), the dividend initiation rates of newly listed firms drop from 8.97% in 1988 to 3.61% in 2001. However, there is a reversal of the trend after 2002. More surprisingly, in the year 2003, 46.57% of the newly listed firms started paying dividends. The initiation rates for firms not paying dividends in the year t-1 have a similar tendency to those of newly listed firms. 3.37% of firms that were listed before 1988 started dividend payments, while in 2001 only 0.45% of firms initiated dividends. Since 2002, however, the initiation rates have started to rise. By the end of 2003, 3.78% of the firms initiated dividend payments, which is the highest for the whole period. This phenomenon will be discussed further in Section 5.

As discussed before, several variables have been shown empirically to play a role in dividend decisions. In this paper, they are defined as follows:

BETA: the slope of the regression of individual stock returns on market returns;

SIZE: the natural log of the product of share prices and shares outstanding;

LR: leverage ratio, defined as the ratio of total liabilities to total assets;

MB: market-to-book ratio, which is the ratio of the market value of total assets to the book value of total assets;

EPR: the earning to price ratio.

SIC: the industrial variables. 1- resource, 2-manufacturing, 3-whole sale and retail, 4-service and 5-public administration and others. Recall that utilities and financial sectors are excluded from the sample.

I use these six determinants of dividend payments to specify the logit model. It is worth mentioning again that dividends seem to have a time trend, as shown by

Fama and French (2001). In different years, investors seem to have different dividend preferences. To capture this time-varying factor and changes in macroeconomic conditions across different years, I do the regression year by year to allow for differential model specifications over the period 1988-2003. The summary statistics of these characteristic variables of dividend initiating firms (D=1) and non-dividend-initiating firms (D=0) for the year 1988 are listed separately in Table 2.

Table 2: Comparison of sample characteristics in different groups (1988)

SIZE=log(data199*data25); LR=data181/data6; MB=(data6-data60+data199*data25)/data6; EPR=data58/data199; SIC is industrial dummies. All these data items are from CompuStat. T-stats are the results of the mean difference tests across two groups. P-values are listed in parentheses.

Characteristic variables[4]	*Correlation with D*	*Mean (D=0)*	*Mean (D=1)*	*T-stats*
BETA	-0.0054	0.6614	0.6445	0.3717 (0.3554)
SIZE	0.1197	4.3007	5.6401	-6.2468 (0.0000)
LR	*-0.0547*	*0.5030*	*0.3606*	*6.1493* *(0.0000)*
MB	*-0.0059*	*1.7926*	*1.7410*	*0.3164* *(0.3759)*
EPR	*0.0261*	*-0.7553*	*-0.0117*	*-1.4069* *(0.1596)*
SIC	*0.0266*	*2.7096*	*2.8491*	*-1.4324* *(0.1596))*

These results are consistent with previous empirical evidence. EPR has a positive effect on dividend initiation, while MB and LR are negatively related to the

[4] I also considered other measures for leverage (debt-to-equity ratio), as well as Tobin's q (market equity to book equity) and earnings (earnings per share). It turns out that these three alternative variables have a lower correlation with D and higher correlations with each other. Furthermore, when they are combined into the specification model, the coefficients of these variables are small and less significant than the variables included. To reduce the multicollinearity problem and guarantee a better specified model, I use the five variables listed in this table.

dividend payments. SIZE is a key risk factor, and has a positive effect on dividend payments. Surprisingly, BETA, MB and EPR is not significantly different between the two groups, which probably results from the presence of some extreme values. However, the logit estimations in section 5 do demonstrate that they have significant effect on the dividend initiation decisions, which is consistent with previous studies. The statistics of SIC are trivial here since SIC is a categorical variable and there are no ordering or ranking involved in its values.

5 Discussion of Results

With the data and the methodology described in Section 3, the propensity-score matching technique is implemented step by step and year by year. I take year 2003 as an example and demonstrate the estimating procedure.

5.1 Logit Regression and Balance Tests

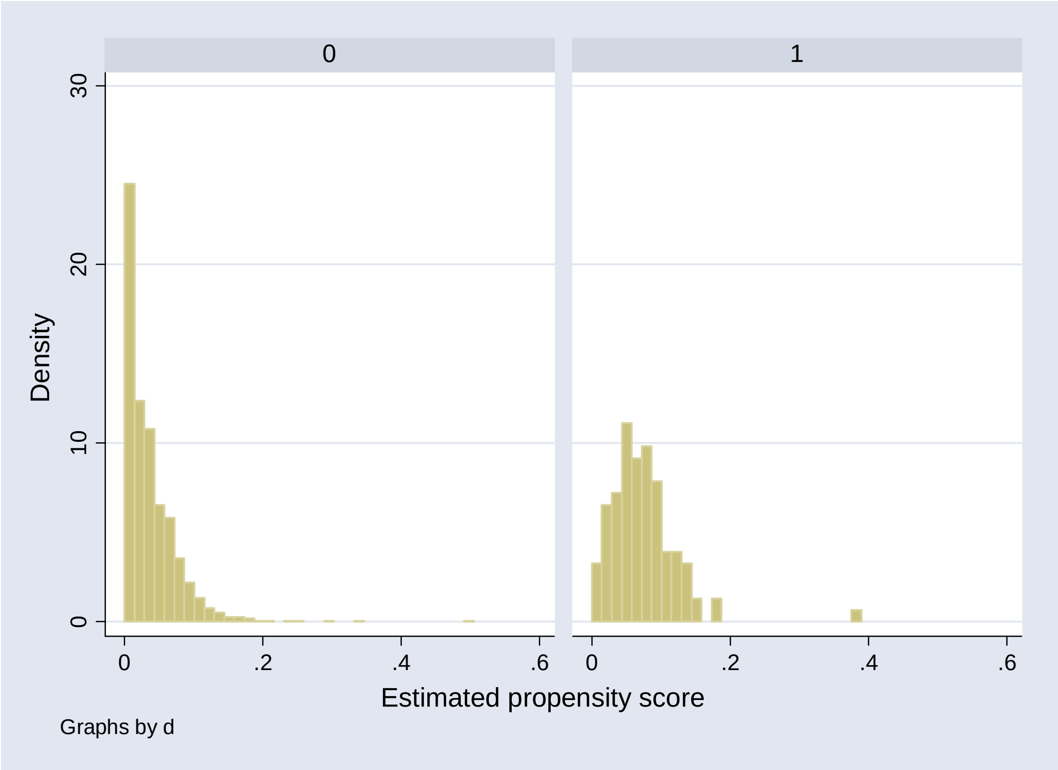
The appropriate model specification for estimating the propensity score should be determined and I start with the simplest linear regression model (1), in which only linear terms of all five characteristic variables are included. The estimated coefficients are quite consistent with the summary statistics discussed in Section 4 except for EPR. The coefficient of EPR is significantly negative, which is consistent with results of previous empirical studies. All of the coefficients are significant at the 5% level except that the coefficient of the market-to-book ratio is significant at the 10% level.

After getting a parsimonious estimation of the propensity scores, I need to test the balance property. The estimated propensity scores are equally split into 5 intervals.

For each interval, the average propensity score of treated and control units should not be significantly different. Finally, I find that 8 blocks to ensure that the mean propensity score is not different for the D group and the ND group. The following balance tests show that with this linear specification, SIZE and EPR are not balanced in blocks 4 and 5. According to Smith and Todd (2003), interaction terms need to be added into the model. Model 2 is specified with one interaction term, SIZE*EPR. The estimated coefficients are listed in Table 3. The coefficients are all significant at the 5% level except that SIC is significant at 10% level. But still, the balance test shows that EPR are not balanced, and this doesn't get any better after the squared term of EPR are added into the model. That leads to Model (3) in which the interaction term BETA*SIZE and EPR*EPR are incorporated. The results show that the balance property is unsatisfied in block 2, 3, 4 and 5, which probably implies that the unbalance is possibly caused by a poor common support.

The distribution of the estimated propensity scores for Model (3) is shown in Figure 1. The panel shows the density of the propensity scores. Most of the ND firms' propensity scores are located between 0 and 0.2. In this histogram, only the units with common support across these two groups are shown. It is obvious that at some specific points, the density of the propensity scores is equal or very close to 0.

Figure 1: The histogram of propensity scores across treatment and comparison groups



To ensure high quality matching, I impose a 6% trimming on the common support and take MB out, which leads to Model (4). In model (4), EPR*LR is significant at the 10% level; all the other coefficients are significant at the 5% level. The resulting log-likelihood is the highest, which implies that EPR*LR has a higher explanatory power than MB. Finally, the balance property is satisfied in all blocks. Thus, Model 4 provides the most reliable propensity score that can be used to match firms in the D and ND groups.

The insignificance of the EPR in Model (4) would not imply that EPR has no explanatory power on dividend initiation decision. Based on our model, three terms cover the information of EPR: SIZE*EPR, EPR*EPR and EPR. Thus, EPR has a multiple nonlinear relationship with D, which can't be revealed by linear regression. Except EPR and BETA*SIZE, all other coefficients are significant either at 5% or at 10% significance level.

The estimated coefficients of the logit model are presented in Table 3. Notice that the significance of the coefficients for the higher-order and interaction terms indicate that those terms reflect the unobserved cofactors other than the five main characteristic variables, and reduce the bias generated by the unobserved confounding factors.

Table 3: Logit results with different specifications (2003)

Model (1) is a parsimonious specification with only linear terms; In Model (2), interaction term of SIZE*EPR is added; Model (3) combines both high order and interaction terms. All the coefficients of these three models are estimated over the estimated common supports. Finally, a 6% trimming level is imposed, which leads to Model (4).

Covariate	Model (1)	Model (2)	Model (3)	Model (4)
Intercept	-4.4744	-4.3228	-3.5945	-3.8656
	(0.4341)	(0.4386)	(0.5322)	(0.5547)
BETA	-0.8751	-0.8119	-1.0707 *	-1.2232
	(0.2202)	(0.2244)	(0.6006)	(0.6057)
SIZE	0.4488	0.4380	0.2731	0.2889
	(0.0628)	(0.0631)	(0.5018)	(0.0988)
LR	-1.5443	-1.7312	-1.7822	-1.8235
	(0.4611)	(0.4778)	(0.5018)	(0.5279)
MB	-0.1867*	-0.2074	-0.1918*	-0.2090*
	(0.0956)	(0.1056)	(0.1012)	(0.1087)
EPR	0.0143	0.0498	-0.4384**	-0.3067**
	(0.0181)	(2.9845)	(0.9680)	(0.3229)
SIC	0.1734*	0.1723*	0.1832*	0.2678
	(0.1009)	(0.1013)	(0.1021)	(0.1094)
*SIZE*EPR*		0.0819	0.7400	0.8147
		(0.0201)	(0.2683)	(0.1848)
*BETA*SIZE*			0.0832*	-0.1021**
			(0.0950)	(0.0986)
*EPR*EPR*			-0.3983**	-0.1830**
			(0.2952)	(0.1052)
R-square	0.0838	0.0931	0.1167	0.1132
Log-likelihood	-416.74	-412.53	-401.79	-365.70

* indicates a 10% significance level, ** indicates insignificance, and all other coefficients are significant at the 5% level. The standard errors are listed in parentheses.

Table 4 presents the result of the balancing test of characteristic variables in Model (4). All the variables are balanced for each block at the 5% significance level. Mean differences are listed in the table and T-statistics are in parentheses. The ideal number of blocks is 4. Thus, the specification of Model (4) could provide a better matching by the estimated propensity scores.

Table 4: Balance test of characteristic variables (Model 4).
This table presents the detailed results of the balance test for Model (4). For each block, the mean difference of pscores and the characteristic variables between the D and ND groups are provided. T-stats of the mean difference tests for all variables are listed in parentheses.

Blocks	Pscore	BETA	SIZE	LR	MB	EPR	EPR
1	-0.45%	0.31	0.78	0.05	0.22	0.17	-0.63*
	(-2.20)	(1.84)	(1.59)	(0.62)	(0.40)	(0.97)	(-2.31)
2	-0.26%	0.08	-0.44	0.03	0.41	-0.02	0.04
	(-1.84)	(0.67)	(1.31)	(0.76)	(1.93)	(-0.25)	(0.22)
3	-0.39%	-0.11	-0.23	-0.03	0.03	-0.03	-0.10
	(-1.79	(-1.41)	(-0.99)	(-0.80)	(0.20)	(-1.39)	(-0.61)
4	-0.33%	-0.09	-0.69	-0.01	-0.59*	-0.03	0.03
	(-1.48)	(-0.76)	(-1.59)	(0.18)	(-2.08)	(-0.44)	(0.11)

5.2 Matching by the estimated propensity score

A detailed discussion about matching methods was provided above in Section 3. The most commonly used method is the single nearest-neighbor matching with replacement.

Table 5 presents the results of the propensity score matching. For each event window, the return effects are all positive and also significant based on T-statistics, which is consistent with previous empirical evidence. However, the ATT is smaller than reported in previous studies, possibly because of the smaller information content of dividend initiation. I leave a detailed discussion of this to Section 5.4. I also divide the whole sample into two groups, one with positive reactions (group P) and the other with negative reactions (group N). Each group shows a much larger average effect compared with the whole sample.

Table 5: Average Effect of Dividend Initiation Announcement

The ATT (PSM) and abnormal returns (event studies) are listed under the mean columns. 0 refers to the effect on declaration day. (-1,1) means the cumulative effect on pre-announcement and announcement day. (-1,1) stands for the cumulative effect over the three-day window around the announcement. Standard errors are in the parentheses of the mean column, while p values are listed in the parentheses of T-stat column.

	Whole Sample		*Positive Reaction(P)*		*Negative Reaction(N)*	
Event	*Mean*	*T-Stat*	*Mean*	*T-stat*	*Mean*	*T-stat*
0	1.49 (0.37)	4.02 (0.00)	5.75 (0.41)	13.99 (0.00)	-5.49 (0.45)	-12.32 (0.00)
(-1,0)	2.33 (0.51)	4.59 (0.00)	6.60 (0.57)	11.59 (0.00)	-4.96 (0.78)	-6.34 (0.00)
(-1,1)	3.49 (0.58)	5.99 (0.00)	7.33 (0.65)	11.35 (0.00)	-3.24 (1.00)	-3.25 (0.00)

Table 6 provides a detailed report on ATTs over the period 1988-2003. The foregoing empirical results generate useful insights, which I discuss next.

Table 6: ATT results across different years

Dividend yield is defined as the ratio of the percentage of the dividend amount on the pre-announcement share price. The bottom line presents the estimated coefficient by regressing ATT on dividend yield. The corresponding standard errors are listed in the parentheses.

Year	*ATT*	*ATT>0*	*ATT<0*	*Dividend yield*
1988	0.43	3.70	-4.61	1.70
1989	1.73	6.45	-4.81	4.53
1990	3.00	6.51	-5.38	9.20
1991	1.76	9.81	-14.99	6.27
1992	0.86	7.66	-7.36	1.50
1993	2.06	4.97	-4.10	0.93
1994	-0.90	5.79	-8.13	1.07
1995	2.12	5.26	-4.25	3.64
1996	1.27	4.11	-5.20	8.66
1997	2.96	6.48	-2.80	3.61
1998	0.87	4.85	-5.09	6.12
1999	2.19	4.44	-2.03	9.78
2000	1.77	3.93	-3.43	5.71
2001	1.09	4.78	-4.07	26.98
2002	2.26	8.68	-7.01	18.13
2003	1.22	5.93	-5.56	7.31
Average	1.54	5.83	-5.55	7.19
Coefficients of DY	0.0589 (0.0197)	0.1009 (0.0203)	-0.0433 (0.0243)	–

5.3 The phenomenon of disappearing dividends and its reversal

Several papers have documented the phenomenon of disappearing dividends. Fama and French (2001) claim that dividends are disappearing in that firms have become less likely to pay dividends. Their explanation is that investors have recognized the tax disadvantage of dividends. Grullon and Michaely (2002) document a decline in both the dividend payout ratio and in the dividend yield. Amihud and Li (2002) claim that there has been a decline since the mid-1970s in the absolute value of the cumulative abnormal return associated with announcements of dividend changes. The dividend response coefficient--the sensitivity of CAR to the magnitude of the dividend change- has declined over time for both dividend increases and decreases. Based on the above evidence, they argue that there has been a decline in the information content of dividend announcements, which has reduced the propensity of firms to pay or increase dividends.

The results in this paper concerning declining dividends are mixed, however. Table 1 demonstrates that dividends are disappearing judging by the proportion of firms initiating dividends; this is confirmed in Table 6 in terms of both dividend yields and stock returns. More interestingly, the initiation rate jumped to 46.57% in 2003. The initiation rates for firms not paying dividends in year t-1 are similar to those for firms newly listed. 3.37% of firms that were listed before 1988 started dividend payments, while in 2001 only 0.55% of firms initiated dividends. Since 2001, the initiation rates have started to increase. By the end of 2003, 3.43% of firms had initiated dividend payments, which is the highest for the whole period 1988-2003. Moreover, Table 6 shows that there was a dramatic jump in dividend yields in 2001 and that stock reactions to the dividend initiation announcement also had a similar turnaround at the beginning of 2000. All this evidence shows that there has been an apparent reversal in the decline in dividend payments, which casts doubt on the phenomenon of disappearing dividends.

The Fama and French (2001) and Amihud and Li (2002) papers do not provide a satisfying explanation for the reversal of the reduction trend in dividend payments. Baker and Wurgler (2002) develop a theory which argues that firms cater to investors' preferences by initiating or omitting dividends. When investors are willing to place a high stock price on dividend payers, firms initiate dividends. Baker and Wurgler focus on the positive relationship between the rates of dividend initiation and omission and on differences in stock prices of payers and non-payers, and explain such differences using four measures of investors' demands for dividend payers. They conclude that the decision to pay dividends is driven by investor demand and that catering is the most natural explanation.

The catering explanation boils down to a benefit-cost argument. Dividends disappear as long as their benefits are less than their costs. Given the cost of a dividend payment, if investors expect the benefits of dividends to increase, the demand for dividend paying firms will increase too. Accordingly, firms will cater to investor preferences and initiate dividends. Conversely, if the benefits are expected to decline, the demand for dividend payers will decrease and firms will cut or omit dividends.

The possible benefit of a dividend payment is its role in signaling a firm's future value and its mitigation of agency costs. These benefits relate to problems of asymmetric information. However, one can argue that this asymmetry may be shrinking as a result of improvements in media technology and the increased number of daily financial analyst reports. In addition, the growing number of institutional investors who tend to be more informed than retail investors, has also reduced the asymmetry of information. As a mechanism for signaling information and mitigating agency costs, dividends may have become less important than before. The benefits of dividends may be getting smaller, while the costs are still high; consequently, dividends are disappearing. However, the question arises, why was there an apparent reversal in the payment of dividends since 2001? I will address this issue in my future work.

5.4 Heterogeneity of Stock Responses

By exploring the data further, I find that 41% of firms that initiated dividends experienced negative ATT at the announcement. Table 5 shows that these effects are significant and that the p values of the t-tests are all below 0.01. In addition, by regressing the ATT of the declaration date on the dividend yields, the estimated coefficient for the whole sample is 0.06, with 0.10 for P group and -0.04 for the N group (see Table 6). All these coefficients are significant at the 5% level. The foregoing implies a positive relation between ATT and the dividend yield for the P group and a negative one for the N group. Such results are reported but not highlighted by Asquith and Mullins (1983), Healy and Palepu (1988) other related papers.

Jin (2000) focuses on this apparent heterogeneity in stock reactions and finds that this phenomenon is not caused by anticipations or confounding events. Both the positive and negative observed reactions are consistent with conventional arguments regarding the information content of dividends and their role in mitigating the agency problem. If the costs of dividend initiations exceed the benefits, a negative reaction ensues, which is consistent with the conventional theory of dividends. The stock responses depend on the net benefits of dividend initiations and differ from firm to firm. This argument is consistent with the cost-benefit argument and the catering theory discussed in Section 5.3.

The previous empirical evidence of positive dividend effects is misleading, since it gives the firms the illusion that the share price would definitely rise after a dividend is initiated. This illusion comes from the signaling hypothesis. The signaling story implies a positive stock response to the dividend initiation, but it is only a necessary condition to signaling, not sufficient. The stock response to the dividend initiation is the result of multiple factors. It could result from the agency cost theory, the tax clientele effect, signaling theory and other unexplored factors. Thus, the dividend initiation effects on stock returns are uncertain and firm-specific and could be explained by the cost-benefit tradeoff model.

5.5 PSM versus Traditional Event Studies

A comparison of the rationales underlying the two methods indicates that propensity score matching outperforms the event study methodology. The results in Table 5 show that the ATT over all the three event windows are very close to the estimated abnormal returns in previous studies. For short-time time horizon, there is little variation in size and other characteristic variables. Thus, in the literature, it is well known that event studies work well for the short time horizon and the similar results show that PSM is a good alternative way for the measurement on the event effects and it is very possible to outperform event studies for long-term horizon measurements since for long term horizon, the other characteristic variables other than BETA would have more effect on the cross section of the stock returns.

6 Conclusion

Prior research has adopted the event study methodology to measure the effect of dividend initiation on stock returns. Although it works well for short-term horizons, the results are seriously questioned by the anomaly literature. Propensity score matching overcomes the curse of dimensionality problem faced by traditional matching techniques and incorporates all the risk factors into the model specification; it is a good way to solve the anomaly problem associated with event studies. I show that the estimated effect of dividend initiations (the ATT) is significantly positive, which is consistent with previous empirical evidence.

While dividends seemed to be disappearing during the period 1988-2000, there has been a dramatic reversal since 2001. This phenomenon may be explained by a dividend benefit-cost argument and by the catering theory. Despite the decreasing information content of dividends, the benefits to investors do not seem to be lower and a turnaround of dividend payments ensues. The apparent heterogeneity in the reactions to dividend initiation observed in this paper can be explained by considering the net effects of the costs and benefits of dividend initiations.

The ATT results are quite close to the estimated abnormal returns in previous studies. It turns out that PSM is a good alternative way for the measurement on the event effects. Moreover, propensity score matching may be employed to measure the after-event long-time performance of stock returns since it is widely recognized that event studies do not perform well over long horizons.

References

Aharony, J. and I. Swary, 1980. Quarterly dividend and earnings announcements and stockholders' returns: An empirical analysis, *The Journal of Finance 31*, *1-12.*

Amihud, Y. and K. Li, 2002. The declining information content of dividend announcements and the effect of institutional holdings, working paper, New York University.

Asquith, P. and D. Mullins, Jr., 1983. The impact of initiating dividend payments on shareholders' wealth, *The Journal of Business* 56, 77–96.

Asquith, P. and D. Mullins, Jr., 1986. Signaling with dividends, stock repurchases and equity issues, *Financial Management*, 27-44.

Baker, M. and J. Wurgler, 2003. A catering theory of dividends, forthcoming in *The Journal of Finance.*

Bhattacharya, Sudipto, 1979. Imperfect information, dividend policy, and 'the bird in the hand' fallacy, *The Bell Journal of Economics 10*, 259-270.

Benesh, G., A. Keown, and J. Pinkerton, 1984. An examination of market reaction to substantial shifts in dividend policy. *The Journal of Financial Research* 7, 131-142.

Charest, Guy, 1978. Dividend information, stock returns and market efficiency - II, *The Journal of Financial Economics* 6, 297-330.

Cheng, Y, 2003, Propensity score matching and the new issues puzzle, *working paper*, *Florida State University*.

Dehejia, R. and S. Wahba, 2002. Propensity score matching methods for non-experimental causal studies, *Review of Economics and Statistics*, 84 (1), 151-161.

Easterbrook, Frank, 1984. Two agency-cost explanations of dividends, *The American Economic Review,* 74(4), 650-659.

Fama, E. and K. French, 1992. The cross-section of expected stock returns, *The Journal of Finance* 47, 427-465.

Fama, E. and K. French, 1993. Common risk factors in the returns on stocks and bonds, *The Journal of Financial Economics* 33, 3-56.

Fama, Eugene F. and Kenneth R. French, 1996. Multifactor explanations of asset price anomalies, *The Journal of Finance,* vol. 51, pp.55-84.

Fama, Eugene F. and Kenneth R. French, 2001. Disappearing dividends: changing firm characteristics or lower propensity to pay? *The Journal of Financial Economics* 60, 3-43.

Fama, Eugene F., L. Fisher, M. Jensen, and R. Roll, 1969. The adjustment of stock prices to new information, *International Economic Review* 10, 1-21.

Healy, Paul M. and Krishna G. Palepu, 1988. Earnings Information Conveyed by Dividend Initiations and Omissions, *The Journal of Financial Economics,* 21(2), 149-176.

Heckman, James, Hidehiko Ichimura, Jeffrey Smith and Petra Todd, 1998. Characterizing selection bias using experimental data, *Econometrica,* 66(5), 1017-1098.

Heckman, James, Hidehiko Ichimura and Petra Todd, 1998. Matching as an econometric evaluation estimator, *Review of Economic Studies,* 65(2), 261-294.

Heckman, James, Hidehiko Ichimura and Petra Todd, 1997. Matching as an econometric evaluation estimator: evidence from evaluating a job training program, *Review of Economic Studies,* 64(4), 605-654.

Heckman, James and Salvador Navarro-Lozano, 2003. Using Matching, Instrumental Variables and Control Functions to Estimate Economic Choice Models, Working paper.

Li, X. and X. Zhao (2003). Is there a seasoned equity offering puzzle? *Working paper, Ohio State University.*

Lintner, John, 1956. Distribution of incomes of corporations among dividends, retained earnings, and taxes, *American Economic Review* 46(2), 97-113.

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

Jin, Z., 2000. On the differential market reaction to dividend initiations.*The Quarterly Review of Economics and Finance* 40, 263-277.

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

Michaely, R., R. Thaler, and K. Womack, 1995. Price reaction to dividend initiations and omissions: overreaction or drift, *The Journal of Finance* 50, 573-608.

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

Miller, Merton and M. Scholes, 1978. Dividends and Taxes. *Journal of Financial Economics,* Dec.1978, 333-364.

Ofter, Aharon R., and Anjan V. Thakor, 1987. A theory of stock price responses to alternative corporate cash disbursement methods: stock repurchases and dividends. *The Journal of Finance* 42, 365-394.

Pettit, R. R., 1972. Dividend announcements, security performance, and capital market efficiency, *The Journal of* Finance 27, 993-1007.

Rimbey, J., and D. Officer, 1992. The market response to subsequent dividend actions of dividend-initiating and omitting firms, *The Quarterly Journal of Business and Economics* 31, 3-20.

Rozeff, Michael S., 1982. Growth, beta and agency costs as determinants of dividend payout ratios, *The Journal of Financial Research* 5, 249-259.

Rosenbaum, Paul R. and Donald B. Rubin, 1983. The central role of the propensity score in observational studies for causal effects, *Biometrika,* 10(1), 41-55.

Rosenbaum, P., D. Rubin,1985. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *American Statistician* 39, 33-38.

Smith, Jeffrey and Petra Todd, 2003. Does matching overcome Lalonde’s critique of non-experimental estimators? Forthcoming in *The Journal of Econometrics.*

Van Eaton, R.D., 1999. Stock price adjustment to the information in dividend changes, *Review of Quantitative and Accounting,* 12, 113-133.

Venkatesh, P.C., 1989. The impact of dividend initiation on the information content of earnings announcements and return volatility, *The Journal of Business* 62, 175-197.