2012

AN INVESTIGATION INTO LONG-RUN ABNORMAL RETURNS USING PROPENSITY SCORE MATCHING

Sunayan Acharya
University of Kentucky, sunayan@gmail.com

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained and attached hereto needed written permission statements(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine).

I hereby grant to The University of Kentucky and its agents the non-exclusive license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless a preapproved embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's dissertation including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Sunayan Acharya, Student

Dr. Bradford Jordan, Major Professor

Dr. Steven Skinner, Director of Graduate Studies

AN INVESTIGATION INTO LONG-RUN ABNORMAL RETURNS USING PROPENSITY SCORE MATCHING

DISSERTATION

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the
Gatton College of Business and Economics
at the University of Kentucky

By

Sunayan Acharya

Lexington, Kentucky

Co-Directors: Dr. Bradford Jordan, Professor of Finance
and Dr. Kristine Hankins, Assistant Professor of Finance

Lexington, Kentucky

2012

ABSTRACT OF DISSERTATION

AN INVESTIGATION INTO LONG-RUN ABNORMAL RETURNS USING PROPENSITY SCORE MATCHING

This is a study in two parts. In part-1, I identify several methods of estimating long-run abnormal returns prevalent in the finance literature and present an alternative using propensity score matching. I first demonstrate the concept with a simple simulation using generated data. I then employ historical returns from CRSP and randomly select events from the dataset using various alternating criteria. I test the efficacy of different methods in terms of type-I and type-II errors in detecting abnormal returns over 12- 36- and 60-month periods. I use various forms of propensity score matching: 1 – 5 Nearest Neighbors in Caliper using distance defined alternatively by Propensity Scores and the Mahalanobis Metric, and Caliper Matching. I show that overall, Propensity Score Matching with two nearest neighbors provides much better performance than traditional methods, especially when the occurence of events is dictated by the presence of certain firm characteristics.

In part-2, I demonstrate an application of Propensity Score Matching in the context of open-market share repurchase announcements. I show that traditional methods are ill-suited for the calculation of long-run abnormal returns following such events. Consequently, I am able to improve upon such methods on two fronts. First, I improve upon traditional matching methods by providing better matches on multiple dimensions and by being able to retain a larger sample of firms from the dataset. Second, I am able to eliminate much of the bias inherent in the Fama-French type methods for this particular application. I show this using simulations on samples based on firms that resemble a typical repurchasing firm. As a result, I obtain a statistically significant 1-, 3-, and 5-year abnormal return of about 2%, 5%, and 10% respectively, which is much lower than what prior literature has shown using traditional methods. Further investigation revealed that much of these returns are unique to small and unprofitable firms.

KEYWORDS: long-run, abnormal, propensity-score, matching, returns

Sunayan Acharya

Student's Signature

29 July, 2012

Date

AN INVESTIGATION INTO LONG-RUN ABNORMAL RETURNS USING PROPENSITY SCORE MATCHING

By

Sunayan Acharya

Dr. Bradford Jordan

Co-Director of Dissertation

Dr. Kristine Hankins

Co-Director of Dissertation

Dr. Steven Skinner

Director of Graduate Studies

29 July, 2012

Date

I dedicate this to the most beautiful thought process in the human mind – Creativity.

ACKNOWLEDGEMENTS

I am grateful to my advisors, Dr. Brad Jordan and Dr. Kristine Hankins for providing me with the necessary direction that has enabled me to successfully complete the work described in this thesis. I am very indebted to Dr. Joe Peek at the Finance department and Dr. Ganpathy Murthy at the department of Physics & Astronomy for being extremely patient and considerate which inadvertantly provided me with much needed moral support. I am also very thankful to the teachers throughout my education who inspired me to be curious and pursue a profession in academia.

I thank my family for love, friends for support, society for its share of resources I have consumed, and humanity for making life worth spending such an enormous amount of effort on. I hope I can repay with interest all that I continue to borrow.

Contents

List of Tables

List of Figures

Chapter 1

Biases in Long-Run Abnormal Returns

1.1 Introduction

An appropriate method to calculate long-run abnormal returns is an unresolved challenge for financial research. In the past decade and a half, several methods have been proposed and tested, ranging from the straightforward control firm approach tested by Barber and Lyon (1997) to more complex methods like the correlation and heteroskedasticity-consistent tests proposed by Jegadeesh and Karceski (2009). However, the methods proposed thus far are imperfect. They haven't comprehensively eliminated various biases that are known to exist in event samples.

In this chapter, we take a step further towards resolving this long-standing issue by investigating the performance of propensity score matching. This method has seen increasing acceptance in corporate finance for calculating abnormal returns in event studies. Essentially, for firms undergoing a certain event, propensity score matching allows us to select control firm(s) for each of these event firms on the basis of a 'propensity score.' In practical applications, the probability of each firm participating in the event is estimated and used as the propensity score. We then allot control firm(s), if any, to each of the event firms on the basis of the proximity of their propensity scores. Finally, abnormal returns are calculated as the difference between the returns of the event and their chosen controls.

While propensity score matching has been used extensively in medical and economic studies evaluating such things as government programs, relatively few studies have been conducted in corporate finance, where the benefits of one corporate policy over another are often extensively discussed. Villagonga (2004) uses two matching estimators to estimate the effect of diversification by single-segment firms on firm value, and finds that contrary to some earlier studies, diversification on average does not destroy value. Hogan and Lewis (2005) document changes in investment behavior of firms that adopted economic profit plans by dividing firms into quartiles based on propensity scores and find that adopters have significantly higher asset turnover compared to non-adopoters. Cooper et al. (2005) examine the effects of mutual fund name changes and find that funds that change their style names earn significantly positive abnormal fund flows in the one year following the name change month. Drucker and Puri (2005) calculate the yield spread difference between loans where the issuer concurrently undewrites the firm's public securities offering and nonconcurrent loans and find that concurrent loans have significantly lower yield spreads. Gande and Puri (2005) estimate ownership-restricted and non-restricted bonds and report a significantly high estimated yield difference between the two cases. Ahn and Walker (2007) find that diversified firms conducting a spin-off exhibit characteristics associated with better corporate governance, as compared to a set of matched peer firms. Aggarwal et al. (2006) compare the governance of foreign firms to matched U.S. firms and find that, on average, foreign firms have worse governance than U.S. firms. Çolak and Whited (2006) estimate whether spin-offs in conglomerate investments increase efficiency and report, using matched samples, that firms that choose to spin-off are different from those that do not, and it is this difference that drives the observed improvements in post spin-off efficiency. Li and Zhao (2006) examine the performance of SEOs and find that matching issuers to non-issuers by propensity scores eliminates any underperformance previously observed. Hellman et al. (2008) find that companies that use relationships developed in the venture capital part of the lending market are able to obtain future loans from these banks at lower yields compared to nonrelationship loans. Marosi and Massoud (2008) find that cross-listed foreign firms that deregister from the U.S. capital market have a lower percentage of U.S. institutional investors compared to non-deregestering firms. Ivashina et al. (2009) investigate the syndicated loan market and find that firms that are the target of attempted takeovers (unsolicited or otherwise) have a higher number of bank-issued loans prior to takeover compared to nontarget firms. Aggarwal et al. (2009) use a firm's probablity of being a foreign firm based on its characteristics as its propensity scores and find that more than 80% of firms across the world invest less in governance compared to a corresponding U.S. firm matched on propensity scores and industry. Masulis et al. (2009) report agency problems at dual-class companies, finding among other results that CEOs at dual-class companies recieve higher compensation and their capital spending contributes less to shareholder wealth compared to single-class companies.

We draw an extensive comparison between propensity score matching and methods that have long been used as a favored tool in calculations of abnormal returns, specifically the control firm approach as described in Lyon and Barber (1997), the RATS method introduced in Ibbotson (1975), and calendar-time portfolios espoused by Fama and French (1993). We show that the use of propensity score matching gives us a much lower incidence of type-I errors in nonrandom samples, where firms are clustered according to some of their underlying characteristics. At the same time, we obtain power comparable to that of the control firm approach. We also follow a unique approach in these tests by considering clustering along multiple characteristics such as size, book-to-market ratio, and momentum simultaneously, where prior research has mostly focused on single characteristic clustering. Given the potential advantages of propensity score matching and the emerging use of this approach in corporate finance, this chapter aims to introduce researchers to the methodology of this technique and to evaluate its performance relative to alternate estimators of long-run abnormal returns in different scenarios.

The rest is organized as follows. Section 2 presents a short history of the development of the propensity score matching method. Section 3 presents a brief numerical simulation highlighting the effectiveness of this technique over simple matching along underlying characteristics. We then apply this technique to long-run abnormal returns in Section 4. Section 5 discusses the results, and Section 6 concludes.

1.2 History

Propensity score analysis developed fairly recently as a statistical tool to calculate treatment effects (the expected gain due to a 'treatment,' for instance, a government policy applicable to a target group) in observational studies. The concept first appeared in a 1983 article by Rosenbaum and Rubin. Contemporary studies with an alternate approach to resolve self-selection biases included Heckman's (1978 and 1979) articles on resolving issues created by dummy endogenous variables using the econometric technique of simultaneous equation modeling. Since then both these approaches have grown substantially to this date and are used extensively to yield unbiased estimates of treatment effects.

The evolution of experimental design accelerated with *The Design of Experiments* by Fisher (1935). Prior to that the common practice in this field was to control for as many factors as possible whenever an experiment was conducted, no doubt influenced by the rather strict requirements of scientific experiments in the laboratory. Fisher, however, noted that such a practice was both practically infeasible and often prohibitively expensive. Instead, he proposed the (at the time) radical idea of not controlling for any confounding factors at all and instead relying on randomization for correct results.[1] Provided that the allocation of units to treatment groups can be randomized (using an appropriate randomization device, for instance, the rolling

[1] Randomized experiments were employed much earlier in the late 1800's by Charles Peirce (1885) among others, but Fisher provided a detailed methodology of the experimental design. In a famous and still highly regarded experiment, eight cups of tea are prepared, four of which had the tea placed first and the other four had milk poured first. A lady was then tested on her claim to be able to determine using taste whether it was tea or milk that was first placed in the cups of tea. The null hypothesis was that the lady didn't possess the requisite skill and conclusions were drawn using Fisher's exact test.

of dice) and the sample size is sufficiently large, it is possible to avoid significant imbalances between the various groups across which the calculation of an effect is desired. Following this, it then became possible to evaluate treatment effects across the randomly assigned groups at various levels of independent (or background) variables of interest, using the technique of matching. For example, the efficacy of a drug can be compared to that of a placebo at various age groups.

Randomization in observational/experimental studies is more of an ideal goal not often encountered in most situations. Heckman (1978, 1979) stressed the importance of addressing sample selection issues in observational/experimental studies, attributing it to two major causes—self-selection by the units being investigated and sample selection decisions by data analysts. These problems cause a bias in treatment effect estimates. For instance, while studying the effect of uninization on wages, the observed wages for union members involves a self-selection bias because the union members found their nonunion alternatives less desirable and chose to be part of the union. Therefore, the wage function estimated on union members would not apply to those from the other group. Thus, a comparison of the wages of union members with those of nonunion workers results in a biased estimate of the effect of unionizing on wages. Since the source of the bias is relegated to an omitted variable, the counterfactual is not observed i.e., we cannot predict what wages the union members would have earned had they not joined a union. This example illustrates the underlying problem in causal inference in observational studies. Once a decision is made regarding the 'treatment status' of the individual unit, in other words which group the individual belongs to, the outcome under the alternative scenario cannot be observed, and hence an accurate judgement of the treatment effect cannot be made. Heckman proposed a model that dealt with this sample selection issue as an omitted variables bias and accounted for this in two steps. The first step estimates the conditional probabilities of the units receiving a particular treatment. The second step adjusts for the bias by adding to the error term its expected value arising from the omitted variable(s). This added term is the inverse Mills ratio calculated from the first step. Thus, the selection bias is explicitly modeled in the outcome regression. This technique contrasted strongly against the statistical approach employed until then that treated the bias as a balancing issue. The central underlying assumption in the latter approach is that any variable(s) that causes a selection bias in the sample is observable. Then, the resolution of this bias amounts to achieving a balanced set of covariates by finding an appropriate control group for the treated units, for instance through a fine-grained matching of the observables. In the above example, if the cause of joining a union were observable then it could be possible to use this as one of the variables to match on and obtain unbiased results.

Matching as a tool in experimental studies has been used as early as in Chapin (1938), Peters (1941), and Greenwood (1945), among several others. Formal evaluation of matching grew with Dorn (1953), Cochran (1953), and Greenberg (1953). In these papers, the authors attempted to lay out a comprehensive discussion on the proper way of selecting control groups and the subsequent matching procedure, as well as comparing the performance of matching against regression in reducing variance in estimated treatment effects[2]—albeit with very simplifying assumptions

[2]The standard response surfaces, i.e. the function relating the independent to the dependent

about the treatment and control populations. Two seminal papers appeared in 1965 that set the stage for much of the later analysis. Billewicz (1965), with the use of an "electronic computer", offered an empirical investigation on the performance of matching against regression on unmatched samples for a variety of situations such as where the underlying relationship between the dependent variable of interest and the observed covariates ranged through linear and parallel, linear and non-parallel, and non-linear but parallel.[3] In this initial study, it was concluded that matching in detecting treatment effects is not superior to regression on random samples, and in fact is mostly marginally inferior. On a similar note, Cochran (1965) performed an analytical study of different matching methods (essentially, different ways of sub-classification) and compared it to a regression-based approach on random samples. Here too the conclusion was that regression seemed superior to matching, although often insubstantially. While matching had continued to be used extensively, due in large part to the simplicity in analysis (especially when the number of covariates aren't many), there was no clear proof that it was superior to regression based methods for observational studies. To this end, Rubin (1973) started empirical investigations to test for the strength of these techniques in reducing bias inherent in the observed variables from non-random samples. This study provided Monte Carlo results for the expected bias in estimated treatment effects assuming different response surfaces and un-symmetric, normally distributed covariate[4] with varying sizes of the control group.

---

variables, considered for the treatment and control groups tended to be linear and parallel, with symmetrical distribution in the covariates. Essentially, this meant both methods had zero bias by design.

[3]For example, $\{Y = X_1$ and $Y = X_2\}$, $\{Y = X_1$ and $Y = 2X_2\}$, and $\{Y = X_1^2$ and $Y = X_2^2\}$ respectively

[4]Only one relevant covariate was assumed for simplicity

Five estimates were considered: the average difference of response in the treatment and control groups (i.e. a simple difference of means), regression-based adjustment of bias assuming a common relationship between the treatment effect and the covariates for both groups, or individually for either group, and matching followed by regression-adjustment.[4] The study found that the matching procedure is especially useful when distributions of the covariate are non-symmetric and/or their variances are unequal between the treatment and control groups. Rubin (1979) extended this study to include two covariates and found stronger support for the use of matching. Thus, when the treatment and control groups are drawn from disparate population groups, matching was found to be very handy in reducing imbalances in the observed covariates. In hindsight, the conclusions drawn from these studies were very significant since in the past two decades there have been a plethora of empirical studies in both social sciences and medical research where such disparities between two groups have been shown to be the norm, rather than the exception.

The concept of propensity scores was introduced in Rosenbaum and Rubin (1983a) as a drastic improvement over the prevailing matching techniques in calculations of the average treatment effect. Most observational studies involve several covariates (hereby X) and matching has exponentially increasing cost as the number of covariates increases—a simple quartile matching across 5 covariates gives rise to 1024 categories! The reason matching is employed is to reduce the bias arising from different values of X in a population having different probabilities of being assigned into

[4]Prior studies had shown that matching followed by regression was almost always superior to matching alone.

a treatment group (the source of the selection bias). Assuming that X contains all observables[5] that are used in the assignment decision, matching ensures that within a sub-class every subject has an equal probability of being treated. This in turn means that both the treatment and control firm are randomly drawn from the same distribution,[6] eliminating any potential bias. Then, for a subject with a given probability of assignment, the assignment decision itself is independent of her characteristics (X).[7] The other essential assumption in matching using sub-classes of X is that for a given x, the treatment effect is independent of the decision to be treated[8]—referred to as the strongly ignorable assumption. This assumption requires that while the treatment itself may lead to a different outcome than that which would be produced by a lack thereof, the decision to be treated shouldn't influence the outcome. The second important result is that if the treatment assignment is strongly ignorable conditional on X, then it is strongly ignorable for the probability of being assigned as well. These theorems then lead to the all-important result that conditional on the probability of being treated (the propensity score) the expected difference in response to treatment in a matched pair equals the average treatment effect. This lets us use matching on propensity score (a univariate criterion) as an alternative to matching on the observables (in general a multivariate criterion), eliminating the need to match on each individual factor that may determine an individual's chances of recieving treatment.[9]

[5]This is implicit, but often not justified in matching studies.

[6]It is important to note that the objective here is to mimic the conditions of a randomized experiment.

[7]Formal proof is provided in Rosenbaum and Rubin (1983a)

[8]As discussed above, violations of this assumption are built into Heckman's model as an omitted variable bias.

[9]Furthermore, if $e(x)$ denotes the propensity score, then if functions f and b exist such that $e(x) = f[b(x)]$, then $b(x)$ (referred to as a balancing score) satisfies the properties mentioned above,

1.3 Propensity Score Matching: A Simple Simulation

In this section, we illustrate the efficacy of the propensity score matching method by using simulated data. As mentioned earlier, the primary purpose of this method is to estimate the treatment effect accurately by utilizing a control group to generate the counterfactual. A simple procedure that captures the essence of the technique is as follows:

We randomly generate a dataset of 10,000 test subjects. A quarter of the units are randomly assigned to the treatment group. This divides the population into two groups: $T_i = 0$ and $T_i = 1$, with T_i denoting the treatment status for unit i. If the treatment and control groups have the same distribution of the underlying characteristics that determine treatment and affect outcomes, then merely averaging the outcomes across the two groups would suffice to generate the treatment effect. The problem arises if these variables have different distributions, the reason for which is made clear below. We let the observed X-variables be normally distributed as follows:

For the untreated group: $X_{j,0} = N(0, 1)$

For the treated group: $X_{j,1} = N(d_j, \sigma_j)$

where, $j : 1 \rightarrow 5$ labels the 5 independent variables, while 0 and 1 in the subscript of X denotes being part of the control and treated groups respectively. Figure 1.1 illustrates the difference in the X-variables between the untreated and the treated

and can be used in lieu of the propensity score.

groups for $d_j = 1$ and $\sigma_j = 1$. The outcome is generated as a continuous variable:

$$Y_i = T_i + x_{1,i} + x_{2,i} + x_{3,i} + x_{4,i} + x_{5,i} + \epsilon_i$$

where, $\epsilon_i\ N(0, \sigma)$. Therefore, the treatment effect by design is 1, as indicated by the presence of the T_i on the right hand side of the above equation. Figure 1.2 shows the approximate difference in the Y-variable distribution between the untreated and treated groups using one covariate (x_1)[10] with $\sigma = 0.05$. This figure signifies the effect of bias created by including the X-variables in both the decision and the outcome equations. The treatment effect, which should be 1, looks to be a lot higher because higher values of X—which lead to higher values of Y—are more likely to be treated. Figure 1.3 illustrates the same effect by plotting the Y-variable against one of the X-variables. Evidently, the appropriate method to use in calculating treatment effect is by looking at the vertical distance between the treatment and the treated outcomes for each X-value and then taking the average for only those X-values that are present in both groups. Such a method, however, is computationally infeasible for multiple X-variables, and approximate methods used instead are unable to exclude the 'tails' effectively from the calculation. Therefore, merely averaging over the two groups will give biased results. Moreover, regression-based methods don't work either, as seen below from attempting an OLS with a treatment dummy:

$$Y_i = \alpha + \beta_0 T_i + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i} + \beta_4 x_{4,i} + \beta_5 x_{5,i} + \epsilon_i$$

For the untreated group:

$$Y_i = \alpha + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i} + \beta_4 x_{4,i} + \beta_5 x_{5,i} + \epsilon_i$$

[10] The difference with all five covariates biased will be much more pronounced.

$$E[Y_i] = E[\alpha + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i} + \beta_4 x_{4,i} + \beta_5 x_{5,i} + \epsilon_i]$$

$$E[Y_i] = \alpha + \beta_1 E[x_{1,i}] + \beta_2 E[x_{2,i}] + \beta_3 E[x_{3,i}] + \beta_4 E[x_{4,i}] + \beta_5 E[x_{5,i}]$$

$$\bar{Y}_{T=0} = \alpha + \beta_1 \bar{x}_{1,T=0} + \beta_2 \bar{x}_{2,T=0} + \beta_3 \bar{x}_{3,T=0} + \beta_4 \bar{x}_{4,T=0} + \beta_5 \bar{x}_{5,T=0}$$

For the treated group:

$$Y_i = \alpha + \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i} + \beta_4 x_{4,i} + \beta_5 x_{5,i} + \epsilon_i$$

$$E[Y_i] = E[\alpha + \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i} + \beta_4 x_{4,i} + \beta_5 x_{5,i} + \epsilon_i]$$

$$E[Y_i] = \alpha + \beta_0 + \beta_1 E[x_{1,i}] + \beta_2 E[x_{2,i}] + \beta_3 E[x_{3,i}] + \beta_4 E[x_{4,i}] + \beta_5 E[x_{5,i}]$$

$$\bar{Y}_{T=1} = \alpha + \beta_0 + \beta_1 \bar{x}_{1,T=1} + \beta_2 \bar{x}_{2,T=1} + \beta_3 \bar{x}_{3,T=1} + \beta_4 \bar{x}_{4,T=1} + \beta_5 \bar{x}_{5,T=1}$$

Subtracting:

$$(\bar{Y}_{T=1} - \bar{Y}_{T=0}) = \beta_0 + \beta_1(\bar{x}_{1,T=1} - \bar{x}_{1,T=0}) + \beta_2(\bar{x}_{2,T=1} - \bar{x}_{2,T=0})$$

$$+ \beta_3(\bar{x}_{3,T=1} - \bar{x}_{3,T=0}) + \beta_4(\bar{x}_{4,T=1} - \bar{x}_{4,T=0}) + \beta_5(\bar{x}_{5,T=1} - \bar{x}_{5,T=0})$$

$$(\bar{Y}_{T=1} - \bar{Y}_{T=0}) = \beta_0 + \beta_1 d_1 + \beta_2 d_2 + \beta_3 d_3 + \beta_4 d_4 + \beta_5 d_5$$

Here, the real difference between the treatment and the control groups is β_0. However, every term to the right of β_0 adds to the bias, with the bias increasing as the the difference between the means in the X-variables of the two groups (d_j) increases.

Since it is practically infeasible to obtain perfectly matched treatment and control units on all the variables, the most common method used to avoid such biases in estimating treatment effects is a form of coarse matching, where the X-variables are divided into quantiles and the treatment effect is calculated as the mean of the mean difference in the outcome variables of the two groups for each quantile. The quantile cut-offs are estimated according to the distribution of the X-variables for the treated units. For instance, using two variables and three quantiles, the effect can be estimated by calculating the quantile means that are then averaged to get the treatment effect, as shown below.

	x_1=*low*	x_1=*mid*	x_1=*high*
x_2=*low*	$\bar{Y}_{1,1} = (\bar{Y}_{T=1} - \bar{Y}_{T=0})$	$\bar{Y}_{1,2} = (\bar{Y}_{T=1} - \bar{Y}_{T=0})$	$\bar{Y}_{1,3} = (\bar{Y}_{T=1} - \bar{Y}_{T=0})$
x_2=*mid*	$\bar{Y}_{2,1} = (\bar{Y}_{T=1} - \bar{Y}_{T=0})$	$\bar{Y}_{2,2} = (\bar{Y}_{T=1} - \bar{Y}_{T=0})$	$\bar{Y}_{2,3} = (\bar{Y}_{T=1} - \bar{Y}_{T=0})$
x_2=*high*	$\bar{Y}_{3,1} = (\bar{Y}_{T=1} - \bar{Y}_{T=0})$	$\bar{Y}_{3,2} = (\bar{Y}_{T=1} - \bar{Y}_{T=0})$	$\bar{Y}_{3,3} = (\bar{Y}_{T=1} - \bar{Y}_{T=0})$

Here, x_1=*low* corresponds to the values of x_1 in the lowest tercile of the treated units' distribution. The other quantiles are defined accordingly. The estimated treatment is then calculated as:

$$Y = \frac{1}{9}\sum_{i,j=1}^{3} \bar{Y}_{i,j}$$

We now show the effect of using propensity score matching on the estimated treatment effect. Propensity scores for treatment effect using 5 covariates is calculated according to the logistic regression:

$$T_i = \gamma + \delta_1 x_{1,i} + \delta_2 x_{2,i} + \delta_3 x_{3,i} + \delta_4 x_{4,i} + \delta_5 x_{5,i} + \nu_i$$

Then, the average treatment effect can be determined by the average difference between the two groups based on matching on propensity scores using any of a variety of methods. In this case we use nearest-neighbor one-to-one matching with replacement, where each treated unit is matched to its closest neighbor (in the control group) in terms of the propensity score. This is the simplest possible matching method. Table 1.1 shows the treatment effect estimates using the steps outlined above. We use five variables to assign treatment and determine outcome (as described above), but let the number of variables used in the estimation vary from one to five. For ordinary matching methods, each variable is divided into two quantiles. Each cell represents the percent bias in the treatment effect estimate. Increasing the difference in the distribution of the X-variables across treatment and control groups (i.e. the induced bias) is expected to increase the outcome bias when using traditional methods. This is shown by the increasing values of the estimates using either a simple average or any of the matching methods. We see from the last row that using all five of the relevant variables in generating propensity scores leads to perfect results with near-zero bias, while the quantile-matching method, which uses 32 bins when employing just two quantiles for each variable, yields considerably poorer results. The bias arising from the treatment and control group distributions differing by half a standard deviation is very high at 92%. However, of greater relevance to practical studies is the comparative advantage of using propensity scores even in the event of not all five of these variables being available or known. As the difference in means of the X-variables between the treatment and the control groups increases, the propensity score method yields an increasingly greater advantage compared to a quantile-matching method. However, we do notice that there is indeed a significant bias in the outcome if all the observables are not accounted for. In the next section, we try to determine if the gains from using propensity scores evidenced here is reflected in the context of observed data, using long-run stock returns.

1.4 Evaluation of Long-Run Returns

As previously noted, propensity score analysis has been used in several studies in corporate finance to estimate long-run returns. In this section, we test this model for misspecifications in simulated event samples. In doing so, we compare the strength of this approach against the alternative of using control firms drawn from size and book-to-market matched samples that have been extensively used in numerous studies over the years.

1.4.1 Literature

Studies of long-run returns following corporate events, seeking to draw inferences about market efficiency, have often documented the presence of abnormalities over periods ranging from one to five years. However, such studies are twin tests of market efficiency as well as the model used in those tests. Several new simulation studies have appeared in the last decade to test these models and improve them to minimize misspecifications. Barber and Lyon (1997) identify three sources of bias often found in test statistics of long-run abnormal returns. These are the *new listing bias*, *rebalancing bias*, and *skewness bias*. These biases arise because of the following reasons:

- Reference portfolios used as indices in calculating abnormal returns typically include stocks that begin trading subsequent to the event month. As has been extensively documented (Ritter (1991)), newly listed firms on an average underperform in the first few months of trading. This gives rise to a positive new listing bias in calculations of abnormal returns involving portfolios with new-listed firms.

- Reference portfolios whose returns are calculated using an equal-weighting scheme are assumed to be rebalanced periodically.[11] On the other hand, no such rebalancing takes place for the sample firms for which the reference portfolios were used as controls to calculate abnormal returns. Since equal weighting can over-estimate the return during a time period due to bid-ask spreads of the component stocks and non-synchronous trading, the corresponding abnormal returns will possibly include a negative rebalancing bias.

- Long-run buy-and-hold abnormal returns calculated using a reference portfolio or a market index are positively skewed. This occurs because as the number of firms in a reference portfolio increases, extreme positive observations become less likely, while individual firms often have high positive returns. As a result, a sample with a positive average abnormal return is more likely to have an extreme positive observation leading to an inflated estimate of the underlying standard deviation. This leads to a deflated t-statistic. On the other hand, a sample with a negative average abnormal return is less likely to have an extreme positive observation leading to a deflated estimate of the underlying standard deviation. This leads to an inflated t-statistic. Thus, in samples with positive abnormal returns, the calculated t-statistic is likely to be lower than the true value, while in samples with negative abnormal returns, the corresponsing t-statistic is likely to be higher than the true value. Thus, we obtain a negative skewness bias in the observed t-statistics.

[11] For instance, if the returns are calculated monthly, then the corresponding portfolios are assumed to be rebalanced monthly as well, with winners sold and losers bought at the end of each month.

Barber and Lyon (1997) found that while t-statistics obtained from cumulative abnormal returns and buy-and-hold abnormal returns are biased, matching sample firms to control firms of similar sizes and book-to-market ratios eliminated these biases and offered considerable improvements in random samples. Kothari and Warner (1997), using various models for calculating abnormal returns such as CAPM and the Fama-French three-factor model, also documented severely biased test statistics. Lyon, Barber, and Tsai (1999) (henceforth LBT) constructed two new statistics—bootstrapped skewness-adjusted t-statistic and p-values based on pseudoportfolios—and found some improvements compared to the control firm approach in some nonrandom samples, but were still unable to avoid all biases. In addition, these test statistics did not perform well in time clustered samples where overlapping returns is an issue. Furthermore, they also find that the calendar time approach advocated by Fama (1998) and used extensively is misspecified in nonrandom samples. Jegadeesh and Karceski (2009) propose a new test of long-run performance by weighting abnormal returns of event firms to account for heteroskedasticity and autocorrelation. Although their proposed test statistics perform better in industry-clustered samples as well as in samples with overlapping returns, they do not perform tests on any other type of clustering.

In spite of the attempts made in the above mentioned studies, the question of which technique best estimates abnormal returns is thus far unresolved. This is significant since calculations of long-run abnormal returns across various studies in corporate finance have been shown to be particularly sensitive to the underlying approach used. The fundamental issue in such studies in estimating abnormal returns is the determination of what 'abnormal' constitutes. Using control firm(s) for a baseline or 'normal' return provides significant ease and is intuitively appealing, and therefore continues to be used in one form or another, with portfolios to single firms playing the role of the control. The basis for the selection of the control has mostly been size and book-to-market, with an increasing number of recent studies incorporating momentum as well. Portfolios are generally constructed by selecting firms from the same quantiles of the relevant characteristics, while single control firms are chosen from the nearest available firms based on those characteristics. As seen in the previous section, this method can lead to significant biases if all relevant characteristics are not matched on, and/or if the differences in the distributions of those characteristics between the treatment and control groups are significantly large. Alternatively, several studies use the calendar time portfolio approach suggested by Fama and French (1999). While this helps significantly in reducing biases in time-clustered samples, the performance in samples clustered on firm characteristics such as size or book-market remains suspect (see Lyon et al. (1999)). In this section we document the existence of such biases and investigate whether the propensity score formulation that was successful in eliminating these biases in a simulated dataset is effective here as well. The propensity score formulation gives us an alternative way of choosing controls, providing much greater flexibility and precision in selecting controls while retaining the intuitive appeal of using controls in defining abnormal returns.

1.4.2 Data and Methodology

We use panel data for all publicly listed firms from NYSE, AMEX, and NASDAQ from 1984 to 2004, keeping only ordinary common shares with share codes 10 and 11. For each firm we obtain financial statement data from the Compustat Annual Database. We use a firm's December filing where available. Otherwise we use the latest filing for that calendar year. A firm is retained in our population for a particular year only if it is on CRSP on December of that year and if we are able to obtain financial statement information for that year from Compustat. We construct the following variables, all dated as of December:

- **Size** - The market value of all outstanding common stock.
- **Book-Market** - The ratio of the book value of common equity to the market value of all outstanding common stocks.
- **Momentum** - The six month buy-and-hold return for the common stock ending in December.
- **Leverage** - The ratio of the total liabilities to the total assets.
- **ROA** - The ratio of the net income to total assets.
- **Cash** - The ratio of total cash and cash equivalents to total assets.
- **TAT** - The ratio of total sales to total assets.

These variables are normalized to have a zero mean and unit standard deviation for ease in interpretation. The summary statistics of these variables are presented in

Table 1.2. Due to the presence of a rather large number of Nasdaq firms, the *Size* variable is highly skewed, with the difference between the minimum and the median values being only about 0.01 standard deviation, while the maximum is 68 standard deviations away from the mean. Since the bottom five size deciles are almost identical, studies employing size-based reference portfolios to calculate long-run returns often use decile breakpoints using NYSE firms. We simulate an event by randomly picking out 200 firm-year observations and assigning them to the treatment group. For each event, we select an event-month of January to be able to use the *Momentum* variable. We then calculate long-run abnormal returns starting from the month of February for 12, 36, and 60 months using the methods described below. We repeat this exercise 2,000 times and compute the percentage of t-statistics from each method that falls in the 5%, 2.5%, or 0.5% tails on left and right side of the corresponding t-distribution. Since the events are generated at random, a well-specified test statistic should fall, for instance, approximately 2.5% of the time in each of the 2.5% tails. The tests used are described below.

1.4.2.1 Propensity Score Matching

A firm's propensity score in any given year is its probability of participating in an event as determined by its underlying characteristics. We assume that the first five normalized variables described earlier cover the factors that can simultaneously influence a firm's participation in an event and affect the firm's future long-run returns. We introduce nonlinearities into the model by including squares of the variables as well. Since a firm's true probability of participation is unknown, we estimate it using a logistic regression:

$$\begin{aligned}Tr_{it} =& \alpha + \beta_1 size_{it} + \beta_1' size_{it}^2 + \beta_2 book - market_{it} + \beta_2' book - market_{it}^2 + \\ & \beta_3 momentum_{it} + \beta_3' momentum_{it}^2 + \beta_4 leverage_{it} + \beta_4' leverage_{it}^2 + \\ & \beta_5 roa_{it} + \beta_5' roa_{it}^2 + \epsilon_{it},\end{aligned}$$

Tr_{it} takes on a value of one if firm i participates in an event in year t. We then assume that the predicted values of Tr_{it} are the true probabilities of the firm i's participation in an event in year t and refer to these as propensity scores. Having obtained the propensity scores for each firm-year, we perform the matching procedure using different methods as listed below. As a preliminary step, we first isolate the set of all control units for each event firm whose propensity scores fall within 0.01 standard deviations of the event firm's propensity score. The standard deviation here is calculated using the predicted values of Tr_{it} from the logistic regression shown above. If this fails to yield even a single match, we expand the radius to 0.02 standard deviations and pick the closest unit as the only control. In this case, if multiple control units have the same propensity score, they are all chosen as possible controls for the corresponding event firm. The value of 0.01 as the initial caliper was empirically determined to be suitable for the current population. The criterion was to ensure sufficient number of matches to work with for further analysis while at the same time to not have too large a caliper as to make the process computationally intractable. It is likely that when the total number of observations in the population is small, the caliper will need to be larger. Control units are replaced after each match and it is therefore possible for two event firms to share one or more controls.

1. **Nearest Neighbor Matching within a Caliper** - Having chosen a set of possible matches for each event firm using the procedure described above, we pick those control units that are closest to it in terms of propensity score alone. The number of control units is allowed to vary from one to five, yielding five different matched sets with one, two, three, four, and five nearest neighbors respectively. Of course, if the initial caliper for an event firm contains less than five control units, those observations are restricted while the remaining are allowed to have upto five control units. Formally, suppose p denotes the estimated propensity scores, and $\{k\}$ is the set of all possible controls within the caliper. Then, unit $\{j_1..j_5\}$ is a control for event firm i if:

    (a) $|p_i - p_{j_1}| < |p_i - p_k|$, where $k \neq j_1$,

    (b) $|p_i - p_{j_1}| \leq |p_i - p_{j_2}| < |p_i - p_k|$, where $k \neq j_1, j_2$,

    (c) $|p_i - p_{j_1}| \leq |p_i - p_{j_2}| \leq |p_i - p_{j_3}| < |p_i - p_k|$, where $k \neq j_1, j_2, j_3$,

    (d) $|p_i - p_{j_1}| \leq |p_i - p_{j_2}| \leq |p_i - p_{j_3}| \leq |p_i - p_{j_4}| < |p_i - p_k|$, where $k \neq j_1, j_2, j_3, j_4$,

    (e) $|p_i - p_{j_1}| \leq |p_i - p_{j_2}| \leq |p_i - p_{j_3}| \leq |p_i - p_{j_4}| \leq |p_i - p_{j_5}| < |p_i - p_k|$, where $k \neq j_1, j_2, j_3, j_4, j_5$

2. **Caliper Matching** - As the limiting case to the previous procedure, we use the entire set of controls for each event firm found in the preliminary step. Here, the number of matches for each event firm ranges from one to many in a matched sample. In other words, the entire set $\{k\}$ is chosen as controls for the event firm. Figure 1.4 shows the number of matches found in random samples of 200

firm-year observations over 2,000 iterations.[12] Of the net 400,000 observations, about 1% result in no matches within the caliper.[13] Of the rest, about 15% of the observations fall in each of the intervals 1-10, 11-20, 21-30, 31-40, and 41-50. Thus, to pick out up to five matches from within the caliper, we start with more than 10 potential matches in more than 85% of the cases. This was one of the criteria used in determining the caliper size.

3. **Nearest Neighbor Mahalanobis Metric Matching within Propensity Score Caliper** - The first procedure uses propensity scores to determine nearest neighbor(s). As an alternative, we consider determining nearest neighbors using the Mahalanobis metric. This requires computing the distance between the event firm and each control firm within the caliper using the Mahalanobis metric. Since we assumed that five firm characteristics are responsible for determining a firm's probability of participating in an event in any given year, we use the same set of five characteristics in estimating the Mahalanobis distance. The distance $d(i, j)$ between event firm i and control unit j is calculated as follows:

$$d(i,j) = \sqrt{(X_i - X_j)^T V^{-1} (X_i - X_j)}$$

Here, X denotes the $5x1$ vector of firm characteristics. V is the covariance matrix for the five variables which is estimated using the entire population. We then pick matches that are closest to each event firm in terms of this distance. Similar to the first procedure, we let the number of control units vary from one

[12]Results are presented in full detail in the next section.
[13]Not shown in the figure.

to five. Analogously, if d denotes the estimated Mahalanobis distance, and $\{k\}$ is the set of all possible controls within the caliper. Then, unit $\{j_1..j_5\}$ is a control for event firm i if:

(a) $|d(i, j_1)| < |d(i, j_k)|$, where $k \neq j_1$,

(b) $|d(i, j_1)| \leq |d(i, j_2)| < |d(i, j_k)|$, where $k \neq j_1, j_2$,

(c) $|d(i, j_1)| \leq |d(i, j_2)| \leq |d(i, j_3)| < |d(i, j_k)|$, where $k \neq j_1, j_2, j_3$,

(d) $|d(i, j_1)| \leq |d(i, j_2)| \leq |d(i, j_3)| \leq |d(i, j_4)| < |d(i, j_k)|$, where $k \neq j_1, j_2, j_3, j_4$,

(e) $|d(i, j_1)| \leq |d(i, j_2)| \leq |d(i, j_3)| \leq |d(i, j_4)| \leq |d(i, j_5)| < |d(i, j_k)|$, where $k \neq j_1, j_2, j_3, j_4, j_5$

Since the Mahalanobis metric uses firm characteristics to measure the distance, the subsequent matching procedure has the possibility of achieving better balancing of covariates between the event firms and their chosen controls. This would be especially useful for highly skewed factors such as *size*.

The abnormal returns and the corresponding paired t-statistics are calculated as follows:

$$AR_i = R_i - \frac{1}{N}\sum_{j=1}^{N} R_j$$
$$t = \frac{mean[AR_i]}{sd[AR_i]/\sqrt{n}}$$

where R_i is the buy-and-hold abnormal return of the event firm over 12, 36, or 60 months, R_j is the buy-and-hold abnormal return of each of N control firm, and n is the number of non-missing observations.

1.4.2.2 Control Firms using Size and Book-Market

- **LBT Control Firm** - In accordance with Lyon et al. (1999), one control firm is chosen for each event firm. We first identify the set of all firms whose size falls within approximately 30%[14] of the the event firm's size for that year. From this set, we chose the firm with the closest book-market ratio to the event firm as the control. This approach was found to be very useful in eliminating the three main sources of biases documented earlier (Lyon and Barber (1997)). It also gives us a yardstick to observe the performance gain or loss in the use of propensity score matching in selecting single control units.

- **Random Control Firms** - We identify the set of all firms in the same size and book-market quintile as the event firm for a particular year. We then pick 5 firms at random as controls. This is similar to the reference portfolios often used in estimating abnormal returns. Like the *LBT Control Firm* above, it gives us a useful benchmark to observe the performance of propensity score matching in selecting five cotrol units.

The abnormal return and t-statistics are calculated using the same procedure as outlined for propensity score matching above.

1.4.2.3 Ibbotson's RATS

We use a modified version of the RATS procedure advocated by Ibbotson (1975) and widely used now (eg. Peyer and Vermaelen (2009)). We first group participating

[14] The actual limits are as follows: We require the control firm size to be at least 25.7% and at most 35% of the event firm size. We use the slightly skewed values to account for the skewness of the distribution of firm size. The specific values were obtained by placing constraints on the log of the firm size instead of the firm size itself.

firms in event-time. For instance, if firm A participated in an event in *Jan, 1987* and firm B in *Jan, 1992*, then:

- event-month *1* would comprise firm A in *Feb, 1987* and firm B in *Feb, 1992*,
- event-month *2* would comprise firm A in *Mar, 1987* and firm B in *Mar, 1992*,
- and so on...

We then run the following regression each event month:

$$R_{it} - R_{ft} = \alpha_{t^*} + \beta_{t^*}(R_{mt} - R_{ft}) + \gamma_{t^*} SMB_t + \delta_{t^*} HML_t + \eta_{t^*} UMD_t + \epsilon_{it},$$

where the dependent and the explanatory variables refer to the participating firm and the usual Fama-French factors respectively in actual month t. Here, t^* indexes the event month. We run this regression for t^* from 1 to 12, 36, and 60 respectively and calculate the cumulative abnormal returns for a time horizon as the sum of the monthly abnormal returns (α_{t^*}). The standard error for each time horizon is calculated as the square root of the sum of squares of the individual standard errors from each event-month regression. Finally, we compute the t-statistic for difference from zero cumulative abnormal return as:

$$t = \frac{AR}{SE},$$

$$AR = \sum_{t^*=1}^{T^*} \alpha_{t^*} \qquad SE = \sqrt{\sum_{t^*=1}^{T^*} se[\alpha_{t^*}]^2}$$

where T^* is one of 12, 36, or 60.

1.4.2.4 Calendar Time Portfolios

Each calendar month in our sample, we select a group of firms that had an event in the past 12, 36, or 60 months (excluding the event month) and form an equal/value-weighted portfolio. Thus, to calculate the monthly abnormal returns:

- The *Jan, 1990* portfolio would include all firms that had an event in the time intervals *Jan–Dec, 1989* for 12 months abnormal returns, *Jan 1987–Dec 1989* for 36 months abnormal returns, and *Jan 1985–Dec 1989* for 60 months abnormal returns respectively.

- and so on for each month from *Jan 1984–Dec 2004*

We then calculate the monthly average abnormal returns using the Fama-French 4-factor model:

$$R_t - R_{ft} = \alpha + \beta(R_{mt} - R_{ft}) + \gamma SMB_t + \delta HML_t + \eta UMD_t + \epsilon_t, \tag{1.1}$$

where R_t is the return on the equal/value-weighted portfolio. The monthly abnormal return is α. The t-statistic is calculated as:

$$t = \frac{\alpha}{sd[\alpha]/\sqrt{T}}$$

where T is the total number of months in our data.

1.5 Results

In their analysis, LBT's control firm method worked well in random samples partly because it corrected for skewness bias. According to them, skewness bias arises because long-run returns of portfolios of multiple stocks are less likely to have extreme positive values than returns of individual stocks. However, there is another factor at play here. In our analysis we assume that if a firm gets delisted prematurely, the investment in that firm is sold at the time of delisting.[15] If this is an event firm, then the observation exits the sample at the time of delisting. The same holds true in case of a solitary control firm. If, however, the delisted firm happens to be one of a portfolio of control firms, then we assume that the proceeds of the sale are reinvested equally into the remaining firms in the portfolio. We do not replace delisted control firms by alternate firm(s) to preserve the quality of the original match.[16] The expected lifetime of a portfolio of multiple stocks is greater than the expected lifetime of an individual stock.[17] Since the median and mean 1-year return of an individual stock are positive, a stock is more likely to have positive returns than negative. Therefore, on an average, the longer the lifetime of a portfolio (with one or more stocks to begin with), the higher the median returns. This can be seen in Figures 1.5, 1.6, and 1.7. Figure 1.5 shows the empirical 12-month buy-and-hold return distribution in two cases: first using 10,000 randomly picked stocks from the CRSP monthly returns database between 1984 and 2004,[18] and then using 5 sets of 10,000 randomly picked stocks and aggregating them into equal-weighted portfolios to provide 10,000 separate portfolios of 5 stocks. These are represented by the solid and dashed lines respectively. As we can see, the dashed line crosses the solid line and stays below

[15] In case of delisting with proceeds, we assume that the investment in the firm is liquidated at the end of the month prior to the delisting. This is done to simplify the analysis by avoiding reinvestments in the middle of a month.

[16] A firm that is chosen non-randomly as a control firm is, by definition, the best available control for the event firm.

[17] where lifetime is defined as the duration for which we continue to earn returns on our investment, which for a portfolio would be for as long as there is at least a single stock left.

[18] we choose this timeframe to maintain consistency with the rest of our analysis.

it. It is therefore true as claimed by LBT that portfolios of multiple stocks are less likely to have extreme positive returns than individual stocks. This is also valid in Figures 1.6 and 1.7 that show the corresponding distributions for 36- and 60-month returns respectively. We also see that in case of 12-month returns, the peak of both distributions occur at about the same point, implying similar median values, but the 5-stock portfolio is a lot shorter on the negative side which should translate into a higher mean value. For the 36- and 60-month returns, similar patterns are exhibited with the additional fact that the peak for the 5-stock portfolio is shifted more to the right, and the entire distribution is right-skewed as well, implying both higher median and mean. The number of stocks in a portfolio used as a reference (or control) to calculate abnormal returns thus plays a significant role in determining the value of the abnormal return. As we increase the number of stocks in the control portfolio, the expected value of the portfolio's average buy-and-hold return for any duration increases. This in turn leads to an increasing negative value for abnormal return.

1.5.1 Random Samples

Results of the specification test on random samples are presented in Table 1.3. The six columns correspond to 0.5%, 2.5%, and 5% on the left and right tails of a t-distribution. The values in each of the cells are the fraction of times (in %) that a t-statistic exceeds the corresponding theoretical value. For instance, the first row of the table shows the values for buy-and-hold abnormal returns calculated using 1 nearest neighbor within caliper using the propensity score distance measure. For this, 0.2% of the 2,000[19] t-statistics fall in the left 0.5% left tail of the corresponding t-distribution, and 0.45% of the t-statistics fall in the right 0.5% tail. Similarly, 2.15% of the t-statistics fall in the left 2.5% tail, and 1.75% fall in the right 2.5% tail. Finally, 4.95% fall in the left 5% tail and 3.9% in the right 5% tail. For a well specified test, we expect the values to be close to the theoretical limits. For the 12-month abnormal returns, using the propensity score distance measure, we see a pattern that was expected. The fraction of t-statistics in the left 5% tail steadily increases from 4.95% with 1 nearest neighbor to 8.1% with 5 nearest neighbors. At the same time, the fraction in the right 5% tail decreases from 3.9% to 2.9%. This is indicative of increasing negative skewness as the number of stocks in the control portfolio increases. Using all the stocks in the caliper (*All in Caliper*), the values in the left and right 5% tails are 9.4% and 2.25% respectively. The median number of stocks in a such portfolio is more than 30, and the values here are therefore indicative of the extent of skewness in the limiting case as the number of stocks in a portfolio keeps increasing. Using the Mahalanobis distance measure to select nearest neighbors, we again see some evidence of skewness. The fraction of t-statistics in the left and right 5% tails goes up from 4.8% and 4.3% respectively with 1 neighbor to 6.35% and 3.6% for 5 neighbors. While there seems to be some improvement in using this measure over propensity score distance, the benefit is eroded away for 36- and 60-month abnormal returns, where we see greater degrees of skewness present.

The RATS procedure is highly skewed to the right with 9.5%, 18.85%, and 28.35% of the t-statistics falling in the right 5% tail for 12-, 36-, and 60-month returns respec-

[19]the number of iterations

tively. Furthermore, using calendar time portfolios with equal-weighted returns gives very highly skewed t-statistics as well. In comparison, the value-weighted portfolio approach performs much better, with the t-statistics for 12- and 36-month returns being well specified. Only the 60-month return exhibits positive skewness, albeit still being much better than when using equal-weighted returns.

The single control firm method advocated by LBT appears to be well specified. Using portfolios of 5 randomly picked control firms in the same size and book-market quintiles as the event firm gives us values similar to the propensity score approach with 5 nearest neighbors. Thus, in random samples, when using a control firm (or portfolio) approach, we see evidence supporting our earlier suggestion that the number of firms in the portfolio plays a significant role in determining the extent of skewness in the test statistic, with the negative skewness bias getting larger as the number of firms in the portfolio increase.

The other side to this degradation in performance from using a large number of firms in the control portfolio is the potential gain in efficiency from an increased number of degrees of freedom which is shown in Table 1.4. Each cell value gives the average number of degrees of freedom used in the calculation of the t-statistic corresponding to the method indicated in the row and the time horizon shown in the column. For the 12-month returns for each iteration, this is the number of observations out of 200 where both the event firm and the control portfolio have non-vanishing returns for 1 year from the beginning of the event. A portfolio is considered to have a return as long as there is at least one stock in the portfolio that has returns present for the entire year. The values for the 36- and 60-month cases are defined analogously.

As expected, as the number of firms in a portfolio increases, the degrees of freedom increases as well, since the probability of at least one surviving firm being present in the portfolio increases with the number of initial firms in the portfolio. We see that for 12-month returns, using the propensity score distance measure leaves us with 83% (166 out of 200) of the initial events that can be used in the calculation of abnormal returns. This can be increased to 90% with 2 or more neighbors but not any further, as seen by the value for *All in Caliper*. However, the difference becomes substantial for 36- and 60-month returns. For 36-month returns, we go from 57.5% to 70.5% to 73.5% with 1, 2, and 3 neighbors respectively with the limit being about 75%. For 60-month returns, we similarly go from 41.5% to 55% to 60% with 1, 2, and 3 neighbors respectively with the limit being about 63%. Thus there is a substantial gain in going from 1 to 2 neighbors, a minor gain from 2 to 3, but not much thereafter. We observe similar numbers using the LBT control firm and the 5 random firms portfolios as well.

We next test for the power of some of these methods in detecting induced abnormal returns. For each of the 2,000 iterations we add an additional 12-month return to each event firm between -10% to +10% in steps of 2%. We then check for the fraction of t-statistics that fall in the left or right 5% tail of the corresponding t-distribution according to the sign of the induced return. The results are shown in Figure 1.8. To maintain legibility, we do not include results from every method shown in Table 1.3, and instead only show values for propensity score matching using the Mahalanobis metric with 1 and 5 nearest neighbors, the LBT control firm, the RATS method, and the calendar time portfolio method with value weighted returns. The improvement in power from using more nearest neighbors becomes apparent as the induced return is increased (in either direction). However, at 5% the difference is only marginal. On the other hand, with a -5% return, there is a larger difference but this can be attributed to the negative skewness of the t-statistic when using 5 nearest neighbors. Using LBT's control firm gives us similar power to that obtained using 1 nearest neighbor. The RATS method, owing to its high positive skewness, yields a very asymmetric curve. The calendar time portfolio approach, using value weighted returns, does provide improvement in power. At 5%(-5%) induced return, it has a power of about 26%(22%), which is higher than the 18%(17%) obtained from using propensity score matching with 1 nearest neighbor.

1.5.2 Nonrandom Samples

We now consider choosing events from clusters of firms with common characteristics. We assume a corporate event in which firms participate with probabilities determined by their characteristics. Suppose, as the values of size, book-market, momentum, leverage, and ROA go up, the probability of a firm participating in the event decreases. To model this, we add these five variables for every firm-year and assign firms to quintiles of this aggregate variable each year. We then randomly choose 200 firms from the lowest quintile as our first clustering criterion. Alternately, we can suppose that as the values of size, book-market, momentum, leverage, and ROA go up, the probability of a firm participating in the event increases instead. In this case, we choose our sample firms from the highest quintile of the aggregate variable as our second clustering criterion.

We use this mode of selection to test the robustness of the propensity score matching method. This is achieved in three ways:

1. The probability of a firm participating in an event depends solely on the sum of its five characteristics as described above. There are about 95,000 observations in our population. Therefore, a firm in the bottom quintile of the aggregate variable defined above in a given year has about a 1%[20] chance of being assigned to the event group in the first case. All other firms in the top 4 quintiles have a probability of 0. Similarly, in the second case, a firm in the top quintile of the aggregate variable in a given year has about a 1% chance of being assigned to the event group. All other firms in the bottom 4 quintiles have a probability of 0. In the logistic regression that estimates the propensity scores, however, we add the five variables above and also their squares. This adds some noise to our formulation.

2. An essential assumption in the use of propensity scores for matching is that all variables that are simultaneously responsible for assigning treatment and affecting outcome should be accounted for. In practical applications it is often difficult to ascertain whether this is being satisfied. We test the robustness of the propensity score method in calculating abnormal returns to the exclusion of relevant factors by omitting two of the variables that we know influence selection. Specifically, we estimate propensity scores by using only size, book-

[20] A quintile holds about 19,000 observations. Since we are picking 200 firm-year observations as our event sample, the probability is approximately 200/19000.

market, and momentum and their squares in the logistic regression:

$$\begin{aligned}Tr_{it} =& \alpha + \beta_1 size_{it} + \beta_1' size_{it}^2 + \beta_2 book - market_{it} + \beta_2' book - market_{it}^2 + \\ & \beta_3 momentum_{it} + \beta_3' momentum_{it}^2 + \epsilon_{it},\end{aligned}$$

3. To ensure that no relevant variables are being missed, a common approach to using propensity scores in research is to use as many variables as possible in the context of the study. It is likely that some of the variables do not play a role in determining participation in an event. We test for the robustness of the propensity score method to inclusion of irrelevant factors by adding in two additional variables, *cash & cash equivalents* and *total asset turnover* and their squares in the logistic regression to estimate propensity scores:

$$\begin{aligned}Tr_{it} =& \alpha + \beta_1 size_{it} + \beta_1' size_{it}^2 + \beta_2 book - market_{it} + \beta_2' book - market_{it}^2 + \\ & \beta_3 momentum_{it} + \beta_3' momentum_{it}^2 + \beta_4 leverage_{it} + \beta_4' leverage_{it}^2 + \\ & \beta_5 roa_{it} + \beta_5' roa_{it}^2 + \beta_6 cash_{it} + \beta_6' cash_{it}^2 + \beta_7 tat_{it} + \beta_7' tat_{it}^2 + \epsilon_{it},\end{aligned}$$

Table 1.5 shows the results of specification tests on samples clustered on the lowest quintile of the constructed variable. The propensity score method is contained in three parts as described above using 3, 5, and 7 variables, presented in that order in the table. We see that when using 3 of the 5 relevant variables, every method using propensity scores fares well for 12-month returns, but gets considerably worse over longer horizons. Using the propensity score metric and 1 nearest neighbor, the fraction of t-statistics in the left and right 5% tails respectively are (4.6%,6.5%) for 12-months, (9.2%,2.35%) for 36-months, and (9.7%,1.7%) for 60-months, with more neighbors yielding higher negative skewness. The Mahalanobis metric gives similar results. Thus, ignoring 2 of the relevant variables clearly biases our results, with the bias getting considerably worse over longer time horizons. With 5 variables, we expect the propensity score method to be reasonably well specified since we are using the very variables that the sample is clustered on. As seen earlier in the results for random samples, we get values quite close to the theoretical limits when using upto 2 nearest neighbors, with both the propensity score metric and the Mahalanobis metric providing similar results. When we add the two irrelevant variables, the Mahalanobis metric does provide some advantage over the propensity score measure. Similar to prior results, we also see that skewness is within acceptable limits as long as the number of nearest neibhbors doesn't exceed 2. With 2 nearest neighbors, the fraction of t-statistics in the left 5% tail with the propensity score measure are 7.15%, 6.95%, and 6.2% for the three time horizons. The corresponding values with the Mahalanobis metric are 6.4%, 5.9%, and 6.1% respectively. LBT's control firm method is negatively skewed with 12.65% in the left 5% tail for 60-month returns. With the RATS procedure, while the fraction of t-stats in the tails doesn't exceed theoretical limits, the total values suggest that the formulation is still misspecified, since the tails get too little. For instance, for the 12-month returns we get (2.55%,4.35%) in both the 5% tails, adding up to 6.9% which trails the theoretical 10%. The equal-weighted calendar time portfolios are still very positively skewed. With the value-weighted portfolios we get much better performance, but the values are still skewed on the negative side and not better than the propensity score method with Mahalanobis metric.

Table 1.6 shows the results of specification tests on samples clustered on the highest quintile of the constructed variable. With 3 variables, as seen before, we get well specified results for 12-month returns, but progressively worse for longer time horizons. With 5 and 7 variables, we get well specified results with 2 nearest neighbors. The values don't get much worse even with more than 2 neighbors. For instance, using the Mahalanobis metric with 7 variables and 2 nearest neighbors, we get (5.05%,5.3%) for 12-months, (4.85%,6.45%) for 36-months, and (3.95%,5.15%) for 60-months in the left and right 5% tails. In this case, however, we don't see much difference between using the propensity score measure and the Mahalanobis measure with 7 variables that we saw in the previous table. Conventional methods have a severe positive skewness. Using LBT's control firm, we get 11.6%, 14.1%, and 12.6% in the right 5% tail for the three time horizons. The corresponding values for the RATS procedure are 17.35%, 42.35%, and 43.8%, respectively, with similar values for the equal-weighted calendar time portfolios. With value-weighted portfolios, we get slightly better results but severely skewed nonetheless, with 8.9%, 10.1%, and 14.7% of the t-statistics in the right 5% tail for 12-, 36-, and 60-month returns.

The results thus far suggest that while the propensity score method with one nearest neighbor yields well specified results, we stand to gain much in terms of efficiency with little lost in terms of bias by going to two nearest neighbors. The Mahalanobis metric may provide better results in some situations compared to the propensity score measure, so in practical applications it may prove useful to apply both and compare the results. We also see that while LBT's control firm method performs identical to the propensity score method with one nearest neighbor in random samples, there is a considerable drop in performance for the former when the samples are drawn from clusters where size and book-market alone are clearly not enough to account for the bias. Ibbotson's RATS method is severely misspecified in all 3 scenarios considered thus far and so are equal-weighted calendar time portfolios. Value-weighted portfolios do yield much better results, but are still very misspecified in clustered samples.

1.5.3 Time-Clustered Samples

Next, we compare the performance of the various methods in time-clustered samples. We randomly select a year and then randomly select 200 firms to assign an event to. The event month is still January. For instance, 200 random firms are assigned an event on *Jan 1994*. This procedure is repeated 2,000 times and the distribution of the t-statistics is recorded as in the prior tables. This way of choosing samples introduces a high degree of cross-sectional correlation. Therefore, we expect that the calendar time portfolios should provide much better results than methods using control portfolios, since the former will discount the effect of time clustering. We should note, however, that in practical applications, events clustered in time are likely to be spread out over a few months instead of all being concentrated in just one month. The method used here, therefore, is an extreme case used to test the robustness of using propensity scores to select matches.

The results are presented in Table 1.7. To estimate propensity scores, we use the 5 variables we originally used with random samples. For 12-month returns, with the propensity score distance, we get 6.4% of the t-statistics in the left 5% tail using 1 nearest neighbor, and 7.35% using 2 nearest neighbors. Mahalanobis metric gives slightly worse results. As the time horizon gets longer, the results get better since the effect of the time clustering gets diluted. Over 36-months, the propensity score distance with 2 nearest neighbors yields 6.2% in the left 5% tail. The corresponding value for 60-months is 5.65%. LBT's control firm performs very similar to the propensity score distance with 1 nearest neighbor with 6.1%, 4.65%, and 4.75% in the left 5% tail for 12-, 36-, and 60-months. The RATS method, operating explicitly in event time, performs very poorly, with almost 80% of the t-statistics falling in the tails corresponding to 10% of the distribution for 12-month returns. With calendar-time portfolios, the equal-weighted approach still remains positively skewed, with over half the t-statistics in the right 5% tail for 60-month returns. The value-weighted portfolios are well-specified for 12- and 36-months, which was expected since the returns are constructed in calendar-time. Only for 60-months do we get a hint of skewness with 7.4% of the t-statistics in the right 5% tail, since the effect of the time-clustering gets weaker and the sample approaches a randomly constructed sample.

In time-clustered samples, therefore, the propensity score distance method with 2 nearest neighbors has a slight negative bias only for 12-month returns. The clustering ceases to matter over longer time horizons. Using 1 nearest neighbor still provides satisfactory results.

1.5.4 Serial-Correlated Samples

We now explicitly introduce serial correlation in our samples. We first randomly select 100 firms in years such that these firms are present in our population for at least one more year. We then select the same firms in the next calendar year to complete our sample. Thus, if firm A was chosen on *Jan 1992*, it is selected again on *Jan 1993*. All other firms are chosen this way as well. This procedure is repeated 2,000 times to complete our test. Since the returns overlap for 36- and 60-month time horizons, any method using control portfolios has a likelihood of being misspecified other than for 12-month returns. With calendar time portfolios, however, we expect well specified results since these portfolios control for serial correlation by counting a firm with multiple occurences only once.

The results are presented in Table 1.8. The propensity score methods with 1 and 2 nearest neighbors are well specified for 12-months. For longer time-horizons, we observe a fat-tails effect. For 36-months with the Mahalanobis metric, we have (6.95%,6.25%) of the t-statistics in the left and right 5% tails with 1 nearest neighbor and (7.65%,6.5%) with 2 nearest neighbors. For 60-months, the corresponding values are (6%,6.25%) with 1 nearest neighbor, and (7.35%,6.85%) with 2 nearest neighbors. As before, with more neighbors, we get higher negative skewness. The propensity score distance gives us slightly worse results. The LBT control firm approach, again, gives similar results to the propensity score method with 1 nearest neighbor. The RATS method, as usual, is severely skewed in the positive direction, as is the calendar time portfolio method with equal-weighting. Value-weighted calendar time portfolios give much better results. We get some evidence of skewness for 60-months, with 7.2% of the t-statistics in the positive 5% tail.

1.6 Conclusions

The tests performed in the previous section suggest that no method works perfectly all the time. There are, however, useful inferences to be drawn from our results. We see that in random samples and in samples where clustering is based on time, LBT's control firm method provides very similar results to those obtained using propensity score matching with one nearest neighbor. The power in random samples is similar as well. Since LBT's method is constrained to use only size and book-market, we observe a drastic performance drop when samples are drawn from clusters based on multiple firm characteristics. It might be possible to posit an alternate control firm scheme, perhaps by replacing size and/or book-market by more suitable factors as demanded by the application, or by adding additional discrete factors such as industry. However, it is not practically feasible to extend this approach to several variables simultaneously. This is the very dilemma that propensity score matching aims to resolve. When using one nearest neighbor, at worst the propensity score matching method performs similar to LBT's control firm method. At best, it is much better at eliminating bias. Adding additional neighbors as controls helps improve efficiency by increasing the numbers of observations of long-run abnormal returns. But this comes at a price of increasing negative skewness. Based on the evidence, 2 nearest neighbors offers the greatest increase in efficiency while at the same time keeping skewness under reasonable limits. When using 2 nearest neighbors, there is some hint of skewness in time-clustered samples over shorter time horizons and samples with high serial correlation over longer time horizons. It should be noted that in such samples, a similar amount of slight skewness also appears in t-statistics from value-weighted calendar time portfolios over 60-month returns.

The fact that the propensity score method doesn't have a one-dimensional series of steps to follow is perhaps what makes it so useful in a wide variety of applications. At the same time, there are several decisions to be made prior to its implementation. The caliper we used was derived from empirical considerations of the distribution of propensity scores. Different applications will in all likelihood have varying propensity score distributions leading to the appropriate caliper size being different as well. The question of which metric to use in measuring distances between treated and control units is a trickier one to answer. In most scenarios considered above, both the propensity score metric and the Mahalanobis metric yielded very similar results. In nonrandom samples while using 7 variables to estimate propensity scores, the Mahalanobis metric generally had a lower incidence of type-I errors than the propensity score distance. We observe slightly better performance by the Mahalanobis metric in serial-correlated samples as well. On the other hand, in time-clustered samples over 12-month horizon, the propensity score distance performs better. In either case, the difference in performance isn't very large. For instance, in Table 1.5 for 36-month returns, while using propensity score matching with 7 variables, the propensity score distance with two nearest neighbors yields 6.95% of the t-statistics in the left 5% tail, while with the Mahalanobis metric we get 5.9%. Meanwhile, in Table 1.7 for 12-month returns, the corresponding numbers are 7.35% vs. 8.15% in favor of the propensity score distance measure. This is typically the extent of the difference between the two metrics. Overall, it appears that the Mahalanobis metric might be at an advantage with samples clustered on firm characteristics because of its potential ability to provide better balance between the event and the control firms in terms of the covariates. This can easily be checked in actual applications.

Ibbotson's RATS method and the equal-weghted calendar time portfolios, both using the Fama-French factors with momentum, are severely skewed. Their use should be avoided in calculations of long-run abnormal returns. Using value-weighted portfolios gives us much better results, with the test statistic being well-specified most of the time in random samples and clustered samples based on time. Its big drawback, however, is in its inability to account for clustering along multiple firm characteristics. Of the methods tested here, therefore, propensity score matching with one or two nearest neighbors is the most promising.

Table 1.1: Mean Bias Percentage in Outcome

This table shows the bias in treatment effect estimates using simple averaging across groups, quantile matching, and propensity score matching. Each value represents the percentage difference from the actual treatment effect as calculated using the method in the corresponding row. The actual treatment effect is 1. The error term in the outcome equation is assumed to be normally distributed with zero mean and a standard deviation of 0.5. Each of the independent variables (X) is normally distributed with a unit standard deviation. d is the difference in means of X-variables between treated and untreated samples. *average* is the difference in average outcome between the treated and untreated populations. *1/5* $\rightarrow$ *5/5* denotes matching using one $\rightarrow$ five of the five independent variables.

d $\longrightarrow$	0.10	0.20	0.30	0.40	0.50
average	49.84	99.22	149.34	200.13	250.24
			Quantile Matching		
1/5 variables	43.72	86.67	130.39	174.92	218.70
2/5 variables	37.02	74.02	111.56	149.53	187.23
3/5 variables	30.76	61.56	92.57	124.39	155.42
4/5 variables	24.58	49.03	73.74	98.76	123.88
5/5 variables	18.22	36.69	55.05	73.56	92.04
			Propensity Score Matching		
1/5 variables	39.98	80.76	119.94	160.33	199.64
2/5 variables	29.76	60.45	90.24	119.86	149.86
3/5 variables	19.70	40.58	59.81	80.49	100.09
4/5 variables	10.18	19.89	29.98	40.56	50.17
5/5 variables	0.01	0.00	0.24	0.26	0.08

Table 1.2: Summary Statistics of Full Sample

This table shows the summary statistics of the variables used in logistic regressions in determining propensity scores of simulated events. All variables are made to have zero mean and unit standard deviation prior to further analysis. *Size* is the market value of common stock, *Book-Market* is the ratio of the market to the book value of common stock, *Momentum* is the six-month buy-and-hold return on the firm's common stock ending in December, *Leverage* is the ratio of total liabilities to total assets, *Return on Assets* is the ratio of the net income to total assets, *Cash & Cash Equivaluents* is the ratio of total cash & cash equivalents to total assets, *Total Asset Turnover* is the ratio of the total annual sales to total assets. All values are obtained from December filings, whenever possible. Otherwise, we use the latest available filing for that calendar year.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	SD
Size	-0.157	-0.154	-0.146	0	-0.103	68.38	1
Book-Market	-0.744	-0.421	-0.182	0	0.156	103.3	1
Momentum	-2.119	-0.514	-0.079	0	0.345	50.33	1
Leverage	-2.106	-0.798	0.026	0	0.748	1.903	1
Return on Assets	-107.8	0.001	0.172	0	0.35	82.98	1
Cash & Cash Equivalents	-0.834	-0.674	-0.433	0	0.274	4.199	1
Total Asset Turnover	-2.278	-0.71	-0.128	0	0.423	67.79	1

Table 1.3: Specification Tests - Random Samples

We randomly select 200 firm-years from our sample and assign each of these an event. For each event, we assign an event month of January. We then calculate long-run abnormal returns using the methods tabulated below. We repeat this exercise 2,000 times and compute the percentage of t-statistics from each method that falls in the 5%, 2.5%, or 0.5% tails on the left and right sides of the corresponding t-distribution. For propensity score matching, we use a caliper of 0.01 standard deviations of the predicted values from the corresponding logistic regression to select our matches from. If this doesn't produce any matches, we use an additional caliper of 0.02 standard deviations from which one match is selected. The variables used to calculate estimated propensity scores are Size, Book-Market, Momentum, Leverage, and Return on Assets, and their squares.

		Theoretical CDF (%)					
		0.5	99.5	2.5	97.5	5	95
12-Month Returns							
		Propensity Score Matching (5 Variables)					
Propensity Score Distance	1 Nearest Neighbor in Caliper	0.20	0.45	2.15	1.75	4.95	3.90
	2 Nearest Neighbors in Caliper	0.65	0.10	3.00	1.45	6.60	3.90
	3 Nearest Neighbors in Caliper	0.90	0.15	3.80	1.25	7.20	3.25
	4 Nearest Neighbors in Caliper	1.20	0.10	4.40	1.15	7.40	2.85
	5 Nearest Neighbors in Caliper	1.20	0.00	4.75	1.20	8.10	2.90
	All in Caliper	1.90	0.00	6.40	0.85	9.40	2.25
Mahalanobis Distance	1 Nearest Neighbor in Caliper	0.40	0.40	2.60	1.80	4.80	4.30
	2 Nearest Neighbors in Caliper	0.75	0.15	3.00	1.60	5.55	4.20

Continued on Next Page…

Table 1.3 – Continued

		Theoretical CDF (%)					
		0.5	99.5	2.5	97.5	5	95
	3 Nearest Neighbors in Caliper	0.85	0.20	3.45	1.25	6.15	3.80
	4 Nearest Neighbors in Caliper	1.25	0.10	3.75	1.10	6.40	3.70
	5 Nearest Neighbors in Caliper	1.30	0.15	3.90	1.20	6.35	3.60
		Conventional Methods					
Control Firm	LBT Control Firm	0.25	0.25	3.05	2.00	5.95	4.55
	5 Random Control Firms	1.10	0.15	4.75	1.15	8.40	3.05
	Ibbotson's RATS	0.05	1.10	0.55	4.75	1.15	9.50
Calendar Time Portfolios	Equal Weghted w/ FF 4-factor	0.05	2.85	0.25	12.75	0.55	22.10
	Value Weighted w/ FF 4-factor	0.45	0.40	1.95	2.20	4.10	4.45
	36-Month Returns						
		Propensity Score Matching (5 Variables)					
Propensity Score Distance	1 Nearest Neighbor in Caliper	0.20	0.35	2.10	1.90	4.80	4.45
	2 Nearest Neighbors in Caliper	0.40	0.25	2.35	1.70	5.10	3.40
	3 Nearest Neighbors in Caliper	0.30	0.00	3.05	0.80	5.65	2.75

Continued on Next Page...

Table 1.3 – Continued

		Theoretical CDF (%)					
		0.5	99.5	2.5	97.5	5	95
	4 Nearest Neighbors in Caliper	0.80	0.00	3.70	0.65	7.05	2.25
	5 Nearest Neighbors in Caliper	1.00	0.00	4.00	0.30	7.45	1.60
	All in Caliper	2.40	0.00	7.15	0.50	11.30	1.75
Mahalanobis Distance	1 Nearest Neighbor in Caliper	0.30	0.30	2.35	1.90	5.15	4.95
	2 Nearest Neighbors in Caliper	0.50	0.10	3.20	1.65	6.60	4.55
	3 Nearest Neighbors in Caliper	0.85	0.25	3.30	1.60	6.20	3.85
	4 Nearest Neighbors in Caliper	1.25	0.25	3.90	1.25	7.10	2.90
	5 Nearest Neighbors in Caliper	1.15	0.15	4.20	1.20	7.15	2.85
		Conventional Methods					
Control Firm	LBT Control Firm	0.40	0.50	2.25	2.15	4.85	4.65
	5 Random Control Firms	1.00	0.05	4.85	0.50	8.55	2.60
	Ibbotson's RATS	0.00	2.45	0.10	10.55	0.20	18.85
Calendar Time Portfolios	Equal Weghted w/ FF 4-factor	0.00	10.90	0.00	30.20	0.05	43.70
	Value Weighted w/ FF 4-factor	0.20	0.85	1.50	3.20	3.55	5.95

Continued on Next Page...

Table 1.3 – Continued

		Theoretical CDF (%)					
		0.5	99.5	2.5	97.5	5	95
60-Month Returns							
		Propensity Score Matching (5 Variables)					
Propensity Score Distance	1 Nearest Neighbor in Caliper	0.15	0.25	2.15	1.80	4.90	4.05
	2 Nearest Neighbors in Caliper	0.45	0.25	3.35	1.40	5.70	3.90
	3 Nearest Neighbors in Caliper	0.70	0.15	3.40	1.10	6.80	2.95
	4 Nearest Neighbors in Caliper	1.10	0.10	4.15	1.30	8.00	2.75
	5 Nearest Neighbors in Caliper	1.40	0.05	4.85	1.00	8.30	2.90
	All in Caliper	3.20	0.00	9.00	0.25	12.85	1.50
Mahalanobis Distance	1 Nearest Neighbor in Caliper	0.25	0.20	1.85	1.95	4.90	4.30
	2 Nearest Neighbors in Caliper	0.50	0.10	3.15	1.65	6.10	4.05
	3 Nearest Neighbors in Caliper	1.00	0.05	3.65	1.05	7.10	3.20
	4 Nearest Neighbors in Caliper	1.10	0.00	4.40	0.95	7.90	3.00
	5 Nearest Neighbors in Caliper	1.10	0.00	4.85	1.15	8.80	3.05
		Conventional Methods					

Continued on Next Page...

Table 1.3 – Continued

		Theoretical CDF (%)					
		0.5	99.5	2.5	97.5	5	95
Control Firm	LBT Control Firm	0.55	0.20	2.70	2.30	5.60	5.00
	5 Random Control Firms	1.45	0.10	4.40	1.30	7.95	2.70
	Ibbotson's RATS	0.05	5.35	0.05	18.00	0.10	28.35
Calendar Time Portfolios	Equal Weghted w/ FF 4-factor	0.00	15.25	0.05	35.90	0.10	49.85
	Value Weighted w/ FF 4-factor	0.50	1.00	2.30	5.55	4.55	9.75

Table 1.4: Degrees of Freedom with Control Portfolios

This table shows the degrees of freedom for t-statistics in specification tests in random samples using various control portfolio approaches. For each of the 2,000 iterations we measure the number of observations (out of 200) where both the event firm and control firm(s) portfolio have long-run returns available on CRSP. We do this for 12-, 36-, and 60-months respectively. For control firm portfolios with more than one firm, the portfolio is considered *active* if at least one stock has returns data available.

		12-Month	36-Month	60-Month
	Propensity Score Matching (5 Variables)			
Propensity	1 Nearest Neighbor in Caliper	166	115	83
Score Distance	2 Nearest Neighbors in Caliper	179	141	110
	3 Nearest Neighbors in Caliper	180	147	120
	4 Nearest Neighbors in Caliper	180	148	123
	5 Nearest Neighbors in Caliper	180	149	124
	All in Caliper	180	149	125
Mahalanobis	1 Nearest Neighbor in Caliper	168	118	87
Distance	2 Nearest Neighbors in Caliper	179	142	112
	3 Nearest Neighbors in Caliper	180	147	120
	4 Nearest Neighbors in Caliper	180	148	123
	5 Nearest Neighbors in Caliper	180	149	124
	Conventional Methods			
Control Firm	LBT Control Firm	168	118	87
	5 Random Control Firms	182	151	126

Table 1.5: Specification Tests - Non-random Samples - I

For each firm-year observation, we define a new variable as the sum of normalized Size, Book-Market, Momentum, Leverage, and Return on Assets. We divide firms into quintiles of this variable each year and then select 200 firm-years from the lowest quintile and assign each of these an event. For each event, we assign an event month of January. This produces non-random samples. We then calculate long-run abnormal returns using the methods tabulated below. We repeat this exercise 2,000 times and compute the percentage of t-statistics from each method that falls in the 5%, 2.5%, or 0.5% tails on the left and right sides of the corresponding t-distribution. For propensity score matching, we use a caliper of 0.01 standard deviations of the predicted values from the corresponding logistic regression to select our matches from. If this doesn't produce any matches, we use an additional caliper of 0.02 standard deviations from which one match is selected. The variables used to calculate estimated propensity scores are Size, Book-Market, Momentum, Leverage, Return on Assets, Cash & Cash-equivalents, and Total Asset Turnover, and their squares. The first three are used in estimating propensity scores with three variables, the first five for five variables, and all for seven variables.

		Theoretical CDF (%)					
		0.5	99.5	2.5	97.5	5	95
	12-Month Returns						
	Propensity Score Matching (3 Variables)						
Propensity Score Distance	1 Nearest Neighbor in Caliper	0.45	0.20	2.10	2.55	4.60	6.50
	2 Nearest Neighbors in Caliper	0.55	0.15	2.45	2.10	4.55	5.45
	3 Nearest Neighbors in Caliper	1.00	0.15	2.65	1.85	4.60	4.90
	4 Nearest Neighbors in Caliper	0.90	0.20	2.90	1.75	5.20	4.55
	5 Nearest Neighbors in Caliper	0.80	0.15	3.05	1.70	5.20	4.80
	All in Caliper	1.15	0.15	3.35	1.75	5.30	4.30

Continued on Next Page...

Table 1.5 – Continued

		Theoretical CDF (%)					
		0.5	99.5	2.5	97.5	5	95
Mahalanobis Distance	1 Nearest Neighbor in Caliper	0.45	0.70	1.70	3.50	3.50	7.30
	2 Nearest Neighbors in Caliper	0.75	0.35	2.35	2.70	4.10	5.45
	3 Nearest Neighbors in Caliper	0.70	0.25	2.50	2.30	4.40	5.20
	4 Nearest Neighbors in Caliper	0.90	0.20	2.55	2.35	4.60	4.95
	5 Nearest Neighbors in Caliper	1.00	0.20	2.80	2.15	5.10	4.60
		Propensity Score Matching (5 Variables)					
Propensity Score Distance	1 Nearest Neighbor in Caliper	0.70	0.10	2.30	1.80	4.70	5.30
	2 Nearest Neighbors in Caliper	0.90	0.25	3.25	1.35	5.50	3.75
	3 Nearest Neighbors in Caliper	0.95	0.05	3.85	1.10	5.95	3.45
	4 Nearest Neighbors in Caliper	1.25	0.20	4.00	1.20	6.10	2.95
	5 Nearest Neighbors in Caliper	1.45	0.10	4.15	1.25	6.60	3.20
	All in Caliper	1.60	0.15	4.25	1.05	6.85	2.35
Mahalanobis Distance	1 Nearest Neighbor in Caliper	0.35	0.25	2.25	2.20	4.60	4.45
	2 Nearest Neighbors in Caliper	0.80	0.20	2.70	1.65	5.05	3.80

Continued on Next Page...

Table 1.5 – Continued

		Theoretical CDF (%)					
		0.5	99.5	2.5	97.5	5	95
	3 Nearest Neighbors in Caliper	0.85	0.25	3.10	1.25	6.05	3.50
	4 Nearest Neighbors in Caliper	1.00	0.10	3.45	1.35	6.05	3.40
	5 Nearest Neighbors in Caliper	1.10	0.20	3.50	1.25	6.50	3.15
		Propensity Score Matching (7 Variables)					
Propensity Score Distance	1 Nearest Neighbor in Caliper	0.50	0.75	2.75	2.30	5.30	4.80
	2 Nearest Neighbors in Caliper	0.70	0.30	4.05	1.90	7.15	4.15
	3 Nearest Neighbors in Caliper	1.05	0.15	4.35	1.80	8.10	3.65
	4 Nearest Neighbors in Caliper	1.05	0.30	4.70	1.50	8.25	3.25
	5 Nearest Neighbors in Caliper	1.45	0.15	5.00	1.55	8.55	3.05
	All in Caliper	1.60	0.10	5.30	1.05	8.80	3.25
Mahalanobis Distance	1 Nearest Neighbor in Caliper	0.35	0.30	2.10	1.80	4.95	4.70
	2 Nearest Neighbors in Caliper	0.55	0.10	3.30	2.05	6.40	4.15
	3 Nearest Neighbors in Caliper	1.15	0.10	4.20	1.80	7.05	3.90
	4 Nearest Neighbors in Caliper	1.55	0.25	4.20	1.50	7.50	3.90

Continued on Next Page…

Table 1.5 – Continued

		Theoretical CDF (%)					
		0.5	99.5	2.5	97.5	5	95
	5 Nearest Neighbors in Caliper	1.65	0.20	4.40	1.70	7.95	3.75
		Conventional Methods					
Control Firm	LBT Control Firm	0.80	0.05	4.40	0.80	7.90	2.85
	5 Random Control Firms	2.65	0.00	5.95	0.45	9.70	1.55
	Ibbotson's RATS	0.10	0.15	1.05	1.90	2.55	4.35
Calendar Time Portfolios	Equal Weghted w/ FF 4-factor	0.00	2.90	0.05	13.50	0.15	23.75
	Value Weighted w/ FF 4-factor	0.65	0.05	3.25	0.75	7.10	2.40
36-Month Returns							
		Propensity Score Matching (3 Variables)					
Propensity Score Distance	1 Nearest Neighbor in Caliper	1.00	0.05	4.95	0.80	9.20	2.35
	2 Nearest Neighbors in Caliper	1.75	0.05	5.85	0.35	11.15	2.05
	3 Nearest Neighbors in Caliper	2.45	0.00	8.00	0.35	11.70	1.90
	4 Nearest Neighbors in Caliper	3.30	0.00	9.25	0.40	14.05	1.60
	5 Nearest Neighbors in Caliper	3.30	0.00	9.80	0.45	14.95	1.45

Continued on Next Page. . .

Table 1.5 – Continued

		Theoretical CDF (%)					
		0.5	99.5	2.5	97.5	5	95
	All in Caliper	5.40	0.00	12.50	0.25	17.60	1.05
Mahalanobis Distance	1 Nearest Neighbor in Caliper	1.15	0.10	4.20	1.00	8.80	2.65
	2 Nearest Neighbors in Caliper	1.70	0.10	6.00	0.90	9.70	1.75
	3 Nearest Neighbors in Caliper	2.50	0.00	7.30	0.75	12.10	1.90
	4 Nearest Neighbors in Caliper	2.70	0.00	8.40	0.70	13.25	2.00
	5 Nearest Neighbors in Caliper	3.05	0.00	9.20	0.45	13.95	1.80
		Propensity Score Matching (5 Variables)					
Propensity Score Distance	1 Nearest Neighbor in Caliper	0.40	0.30	1.90	1.60	4.25	3.75
	2 Nearest Neighbors in Caliper	0.60	0.30	2.45	1.30	6.00	3.00
	3 Nearest Neighbors in Caliper	0.70	0.15	3.30	0.75	7.80	3.05
	4 Nearest Neighbors in Caliper	1.20	0.15	3.90	1.00	7.70	2.65
	5 Nearest Neighbors in Caliper	1.50	0.10	4.45	0.80	8.55	2.80
	All in Caliper	2.25	0.10	5.70	0.65	9.35	1.85
Mahalanobis Distance	1 Nearest Neighbor in Caliper	0.20	0.15	2.05	1.60	4.35	3.40

Continued on Next Page...

Table 1.5 – Continued

		Theoretical CDF (%)					
		0.5	99.5	2.5	97.5	5	95
	2 Nearest Neighbors in Caliper	0.50	0.05	3.05	0.90	6.35	3.05
	3 Nearest Neighbors in Caliper	0.85	0.10	3.40	1.20	7.05	2.65
	4 Nearest Neighbors in Caliper	0.80	0.05	3.90	0.95	7.10	2.50
	5 Nearest Neighbors in Caliper	1.15	0.05	4.10	0.90	7.80	2.40
		Propensity Score Matching (7 Variables)					
Propensity Score Distance	1 Nearest Neighbor in Caliper	0.15	0.30	2.70	2.05	4.80	4.85
	2 Nearest Neighbors in Caliper	0.65	0.40	3.55	1.75	6.95	3.90
	3 Nearest Neighbors in Caliper	1.05	0.25	4.50	1.25	8.30	3.30
	4 Nearest Neighbors in Caliper	1.30	0.10	5.50	1.55	8.80	3.00
	5 Nearest Neighbors in Caliper	1.70	0.20	6.00	1.55	9.85	3.05
	All in Caliper	2.25	0.05	6.85	0.90	10.65	2.55
Mahalanobis Distance	1 Nearest Neighbor in Caliper	0.20	0.40	2.05	2.15	4.40	5.10
	2 Nearest Neighbors in Caliper	0.70	0.30	3.00	1.70	5.90	4.50
	3 Nearest Neighbors in Caliper	0.85	0.20	3.70	1.50	6.85	3.45

Continued on Next Page. . .

Table 1.5 – Continued

		Theoretical CDF (%)					
		0.5	99.5	2.5	97.5	5	95
	4 Nearest Neighbors in Caliper	1.00	0.10	4.40	1.60	7.70	3.55
	5 Nearest Neighbors in Caliper	1.20	0.15	4.65	1.30	7.95	3.30
		Conventional Methods					
Control Firm	LBT Control Firm	1.10	0.00	5.10	0.55	9.05	2.15
	5 Random Control Firms	3.35	0.00	10.60	0.20	15.55	0.95
	Ibbotson's RATS	0.15	0.55	0.75	2.65	1.60	5.70
Calendar Time Portfolios	Equal Weghted w/ FF 4-factor	0.00	13.45	0.00	39.15	0.00	54.70
	Value Weighted w/ FF 4-factor	0.65	0.10	3.25	0.90	6.70	2.40
	60-Month Returns						
		Propensity Score Matching (3 Variables)					
Propensity Score Distance	1 Nearest Neighbor in Caliper	1.20	0.10	5.10	0.70	9.70	1.70
	2 Nearest Neighbors in Caliper	1.65	0.05	7.20	0.40	12.55	1.55
	3 Nearest Neighbors in Caliper	3.00	0.05	9.40	0.35	14.60	1.00
	4 Nearest Neighbors in Caliper	3.65	0.05	9.80	0.25	15.55	0.80

Continued on Next Page...

Table 1.5 – Continued

		Theoretical CDF (%)					
		0.5	99.5	2.5	97.5	5	95
	5 Nearest Neighbors in Caliper	4.55	0.05	10.80	0.35	16.80	0.85
	All in Caliper	6.50	0.05	15.05	0.25	20.80	0.45
Mahalanobis Distance	1 Nearest Neighbor in Caliper	0.75	0.05	4.65	0.60	9.25	2.00
	2 Nearest Neighbors in Caliper	1.50	0.05	6.45	0.15	12.30	1.35
	3 Nearest Neighbors in Caliper	2.75	0.05	8.65	0.35	14.20	1.00
	4 Nearest Neighbors in Caliper	3.55	0.05	9.90	0.20	15.65	0.85
	5 Nearest Neighbors in Caliper	4.35	0.10	11.15	0.35	16.30	0.80
		Propensity Score Matching (5 Variables)					
Propensity Score Distance	1 Nearest Neighbor in Caliper	0.25	0.30	2.35	2.20	4.65	4.75
	2 Nearest Neighbors in Caliper	0.60	0.20	2.50	1.20	5.85	3.65
	3 Nearest Neighbors in Caliper	0.80	0.15	3.25	1.10	6.30	2.80
	4 Nearest Neighbors in Caliper	1.05	0.10	3.90	1.00	8.15	2.40
	5 Nearest Neighbors in Caliper	1.15	0.10	4.40	0.85	8.70	1.85
	All in Caliper	1.95	0.05	6.15	0.60	10.00	1.20

Continued on Next Page...

Table 1.5 – Continued

		Theoretical CDF (%)					
		0.5	99.5	2.5	97.5	5	95
Mahalanobis Distance	1 Nearest Neighbor in Caliper	0.35	0.15	2.50	1.95	5.10	4.30
	2 Nearest Neighbors in Caliper	0.55	0.10	3.30	1.20	6.00	3.40
	3 Nearest Neighbors in Caliper	0.90	0.15	3.55	1.50	7.10	3.05
	4 Nearest Neighbors in Caliper	1.15	0.05	3.75	0.90	7.55	2.50
	5 Nearest Neighbors in Caliper	1.15	0.10	4.75	1.00	8.05	2.45
		Propensity Score Matching (7 Variables)					
Propensity Score Distance	1 Nearest Neighbor in Caliper	0.10	0.15	2.25	1.50	6.10	3.80
	2 Nearest Neighbors in Caliper	0.50	0.10	3.35	1.35	6.20	2.95
	3 Nearest Neighbors in Caliper	1.25	0.10	4.05	1.00	7.65	2.65
	4 Nearest Neighbors in Caliper	1.30	0.15	4.75	0.85	8.50	2.25
	5 Nearest Neighbors in Caliper	1.50	0.10	5.50	0.65	9.40	2.25
	All in Caliper	2.50	0.00	7.15	0.35	11.55	1.70
Mahalanobis Distance	1 Nearest Neighbor in Caliper	0.15	0.25	1.75	1.40	4.05	3.85
	2 Nearest Neighbors in Caliper	0.30	0.25	2.85	1.70	6.10	3.50

Continued on Next Page...

Table 1.5 – Continued

		Theoretical CDF (%)					
		0.5	99.5	2.5	97.5	5	95
	3 Nearest Neighbors in Caliper	0.60	0.05	3.70	1.05	6.70	3.25
	4 Nearest Neighbors in Caliper	1.05	0.05	4.15	0.90	7.20	2.90
	5 Nearest Neighbors in Caliper	1.30	0.10	4.60	0.95	8.15	2.80
		Conventional Methods					
Control Firm	LBT Control Firm	1.70	0.00	7.15	0.50	12.65	1.55
	5 Random Control Firms	4.35	0.00	12.15	0.10	18.05	0.45
	Ibbotson's RATS	0.05	0.50	0.55	2.95	1.05	6.60
Calendar Time Portfolios	Equal Weghted w/ FF 4-factor	0.00	17.35	0.00	44.70	0.00	60.40
	Value Weighted w/ FF 4-factor	0.85	0.10	3.75	1.25	7.75	2.65

Table 1.6: Specification Tests - Non-random Samples - II

For each firm-year observation, we define a new variable as the sum of normalized Size, Book-Market, Momentum, Leverage, and Return on Assets. We divide firms into quintiles of this variable each year and then select 200 firm-years from the highest quintile and assign each of these an event. For each event, we assign an event month of January. This produces non-random samples. We then calculate long-run abnormal returns using the methods tabulated below. We repeat this exercise 2,000 times and compute the percentage of t-statistics from each method that falls in the 5%, 2.5%, or 0.5% tails on the left and right sides of the corresponding t-distribution. For propensity score matching, we use a caliper of 0.01 standard deviations of the predicted values from the corresponding logistic regression to select our matches from. If this doesn't produce any matches, we use an additional caliper of 0.02 standard deviations from which one match is selected. The variables used to calculate estimated propensity scores are Size, Book-Market, Momentum, Leverage, Return on Assets, Cash & Cash-equivalents, and Total Asset Turnover, and their squares. The first three are used in estimating propensity scores with three variables, the first five for five variables, and all for seven variables.

		Theoretical CDF (%)					
		0.5	99.5	2.5	97.5	5	95
12-Month Returns							
Propensity Score Matching (3 Variables)							
Propensity Score Distance	1 Nearest Neighbor in Caliper	0.05	0.90	0.75	4.50	2.85	7.70
	2 Nearest Neighbors in Caliper	0.10	0.45	1.70	3.10	4.30	6.40
	3 Nearest Neighbors in Caliper	0.30	0.35	2.45	2.35	4.80	5.65
	4 Nearest Neighbors in Caliper	0.35	0.35	2.55	2.30	5.20	5.65
	5 Nearest Neighbors in Caliper	0.60	0.30	2.50	2.10	5.20	4.85
	All in Caliper	0.85	0.10	3.30	1.70	5.60	4.40

Continued on Next Page...

Table 1.6 – Continued

		Theoretical CDF (%)					
		0.5	99.5	2.5	97.5	5	95
Mahalanobis Distance	1 Nearest Neighbor in Caliper	0.25	0.40	1.95	2.80	5.15	4.95
	2 Nearest Neighbors in Caliper	0.55	0.25	2.60	1.55	5.80	4.15
	3 Nearest Neighbors in Caliper	0.65	0.25	2.80	1.45	6.00	3.45
	4 Nearest Neighbors in Caliper	0.65	0.20	3.40	1.30	6.10	3.45
	5 Nearest Neighbors in Caliper	0.70	0.20	3.25	1.55	5.95	3.15
		Propensity Score Matching (5 Variables)					
Propensity Score Distance	1 Nearest Neighbor in Caliper	0.20	0.30	2.40	2.90	4.40	5.40
	2 Nearest Neighbors in Caliper	0.60	0.15	3.00	2.15	5.00	5.10
	3 Nearest Neighbors in Caliper	0.55	0.00	3.05	1.70	5.50	4.60
	4 Nearest Neighbors in Caliper	0.65	0.15	3.30	1.95	6.00	4.25
	5 Nearest Neighbors in Caliper	0.85	0.20	3.50	1.70	6.10	3.75
	All in Caliper	1.10	0.20	3.70	1.25	6.00	3.40
Mahalanobis Distance	1 Nearest Neighbor in Caliper	0.35	0.55	2.30	2.25	4.65	4.90
	2 Nearest Neighbors in Caliper	0.60	0.25	2.55	1.70	5.40	4.40

Continued on Next Page...

Table 1.6 – Continued

		Theoretical CDF (%)					
		0.5	99.5	2.5	97.5	5	95
	3 Nearest Neighbors in Caliper	0.45	0.25	3.30	1.25	5.55	3.85
	4 Nearest Neighbors in Caliper	0.65	0.20	3.15	1.45	5.80	3.65
	5 Nearest Neighbors in Caliper	0.80	0.15	3.05	1.20	6.25	3.55
		Propensity Score Matching (7 Variables)					
Propensity Score Distance	1 Nearest Neighbor in Caliper	0.35	0.40	1.60	2.90	3.45	6.30
	2 Nearest Neighbors in Caliper	0.35	0.30	2.25	2.55	4.25	6.10
	3 Nearest Neighbors in Caliper	0.55	0.25	2.40	2.25	4.85	5.30
	4 Nearest Neighbors in Caliper	0.55	0.15	2.90	2.05	5.60	5.10
	5 Nearest Neighbors in Caliper	0.55	0.15	2.80	2.05	5.50	4.55
	All in Caliper	0.90	0.10	2.70	1.35	5.75	4.25
Mahalanobis Distance	1 Nearest Neighbor in Caliper	0.55	0.45	2.00	2.85	3.95	5.65
	2 Nearest Neighbors in Caliper	0.45	0.45	2.15	2.15	5.05	5.30
	3 Nearest Neighbors in Caliper	0.65	0.30	2.60	1.95	5.05	5.00
	4 Nearest Neighbors in Caliper	0.75	0.30	2.70	2.05	5.45	4.90

Continued on Next Page. . .

Table 1.6 – Continued

		Theoretical CDF (%)					
		0.5	99.5	2.5	97.5	5	95
	5 Nearest Neighbors in Caliper	0.80	0.30	2.95	1.95	5.05	4.75
		Conventional Methods					
Control Firm	LBT Control Firm	0.00	2.15	0.60	6.65	1.60	11.60
	5 Random Control Firms	0.05	1.00	0.65	6.80	1.40	13.35
	Ibbotson's RATS	0.00	2.65	0.20	9.75	0.65	17.35
Calendar Time Portfolios	Equal Weghted w/ FF 4-factor	0.00	11.75	0.00	32.90	0.00	47.40
	Value Weighted w/ FF 4-factor	0.05	0.70	0.65	3.75	1.70	8.90
36-Month Returns							
		Propensity Score Matching (3 Variables)					
Propensity Score Distance	1 Nearest Neighbor in Caliper	0.15	1.65	0.50	6.45	1.65	11.25
	2 Nearest Neighbors in Caliper	0.05	1.25	0.85	5.25	1.85	10.30
	3 Nearest Neighbors in Caliper	0.00	0.90	0.80	5.25	2.30	9.50
	4 Nearest Neighbors in Caliper	0.00	1.00	1.05	4.60	2.70	8.95
	5 Nearest Neighbors in Caliper	0.10	0.65	1.00	4.50	2.65	8.35

Continued on Next Page…

Table 1.6 – Continued

		Theoretical CDF (%)					
		0.5	99.5	2.5	97.5	5	95
	All in Caliper	0.40	0.45	1.65	3.35	3.25	7.30
Mahalanobis Distance	1 Nearest Neighbor in Caliper	0.00	1.35	0.50	4.95	1.95	8.80
	2 Nearest Neighbors in Caliper	0.05	0.70	0.85	4.70	2.00	8.90
	3 Nearest Neighbors in Caliper	0.10	0.45	0.85	4.15	2.00	7.30
	4 Nearest Neighbors in Caliper	0.15	0.60	1.00	4.30	2.60	8.10
	5 Nearest Neighbors in Caliper	0.10	0.55	1.00	3.95	2.65	6.85
		Propensity Score Matching (5 Variables)					
Propensity Score Distance	1 Nearest Neighbor in Caliper	0.20	0.35	2.00	2.20	4.20	5.75
	2 Nearest Neighbors in Caliper	0.85	0.15	3.05	1.60	5.35	4.05
	3 Nearest Neighbors in Caliper	0.90	0.10	3.55	1.60	6.30	3.95
	4 Nearest Neighbors in Caliper	1.40	0.10	4.25	1.70	6.75	3.65
	5 Nearest Neighbors in Caliper	1.90	0.10	4.50	1.50	7.50	3.30
	All in Caliper	1.95	0.05	5.00	0.95	8.40	2.70
Mahalanobis Distance	1 Nearest Neighbor in Caliper	0.35	0.15	2.25	2.40	4.85	5.70

Continued on Next Page...

Table 1.6 – Continued

		Theoretical CDF (%)					
		0.5	99.5	2.5	97.5	5	95
	2 Nearest Neighbors in Caliper	0.45	0.30	3.10	1.90	6.05	4.85
	3 Nearest Neighbors in Caliper	0.70	0.00	3.25	1.60	6.40	4.30
	4 Nearest Neighbors in Caliper	0.95	0.00	3.70	1.25	6.35	4.00
	5 Nearest Neighbors in Caliper	1.10	0.10	3.75	1.15	6.65	4.20
		Propensity Score Matching (7 Variables)					
Propensity Score Distance	1 Nearest Neighbor in Caliper	0.20	0.20	1.70	1.90	4.30	4.65
	2 Nearest Neighbors in Caliper	0.65	0.20	2.60	1.95	5.30	4.80
	3 Nearest Neighbors in Caliper	0.40	0.10	3.20	1.70	5.85	4.05
	4 Nearest Neighbors in Caliper	0.75	0.05	3.95	1.55	6.85	4.35
	5 Nearest Neighbors in Caliper	0.75	0.05	4.30	1.45	7.15	3.95
	All in Caliper	0.95	0.00	4.90	1.30	7.80	3.50
Mahalanobis Distance	1 Nearest Neighbor in Caliper	0.45	0.35	2.20	2.90	4.05	6.65
	2 Nearest Neighbors in Caliper	0.60	0.20	2.50	2.60	4.85	6.45
	3 Nearest Neighbors in Caliper	0.55	0.15	2.55	2.40	5.15	5.55

Continued on Next Page...

Table 1.6 – Continued

		Theoretical CDF (%)					
		0.5	99.5	2.5	97.5	5	95
	4 Nearest Neighbors in Caliper	0.65	0.10	2.90	2.00	5.35	5.70
	5 Nearest Neighbors in Caliper	0.70	0.00	3.20	2.00	6.05	5.40
		Conventional Methods					
Control Firm	LBT Control Firm	0.00	1.95	0.40	7.95	1.10	14.10
	5 Random Control Firms	0.10	0.75	1.45	4.25	2.95	8.20
	Ibbotson's RATS	0.00	11.95	0.05	30.15	0.10	42.35
Calendar Time Portfolios	Equal Weghted w/ FF 4-factor	0.00	18.80	0.00	42.40	0.00	56.65
	Value Weighted w/ FF 4-factor	0.00	1.05	0.30	4.90	1.05	10.10
60-Month Returns							
		Propensity Score Matching (3 Variables)					
Propensity Score Distance	1 Nearest Neighbor in Caliper	0.10	1.00	0.60	5.20	1.80	8.85
	2 Nearest Neighbors in Caliper	0.20	0.55	0.75	3.65	2.45	8.05
	3 Nearest Neighbors in Caliper	0.10	0.65	1.05	3.05	2.95	6.65
	4 Nearest Neighbors in Caliper	0.15	0.45	0.95	2.65	3.10	5.75

Continued on Next Page. . .

Table 1.6 – Continued

		Theoretical CDF (%)					
		0.5	99.5	2.5	97.5	5	95
	5 Nearest Neighbors in Caliper	0.15	0.40	1.45	2.65	3.85	5.45
	All in Caliper	0.95	0.25	3.20	1.90	5.90	4.70
Mahalanobis Distance	1 Nearest Neighbor in Caliper	0.00	1.35	0.50	4.60	1.70	8.75
	2 Nearest Neighbors in Caliper	0.00	0.90	0.95	4.90	2.15	9.05
	3 Nearest Neighbors in Caliper	0.05	0.75	0.80	3.65	2.30	8.40
	4 Nearest Neighbors in Caliper	0.10	0.70	0.85	3.35	2.50	7.45
	5 Nearest Neighbors in Caliper	0.25	0.65	1.15	3.65	2.60	6.85
		Propensity Score Matching (5 Variables)					
Propensity Score Distance	1 Nearest Neighbor in Caliper	0.20	0.35	2.05	2.30	4.25	5.05
	2 Nearest Neighbors in Caliper	0.40	0.15	2.10	1.85	4.05	4.50
	3 Nearest Neighbors in Caliper	0.95	0.10	2.65	1.60	5.50	3.80
	4 Nearest Neighbors in Caliper	0.95	0.25	3.20	1.20	5.80	3.30
	5 Nearest Neighbors in Caliper	1.00	0.20	3.20	1.20	6.50	3.70
	All in Caliper	1.65	0.20	4.80	0.75	8.15	2.60

Continued on Next Page...

Table 1.6 – Continued

		Theoretical CDF (%)					
		0.5	99.5	2.5	97.5	5	95
Mahalanobis Distance	1 Nearest Neighbor in Caliper	0.30	0.50	2.05	2.45	3.65	4.95
	2 Nearest Neighbors in Caliper	0.15	0.45	1.75	1.90	4.50	4.25
	3 Nearest Neighbors in Caliper	0.30	0.25	1.80	1.65	4.05	4.65
	4 Nearest Neighbors in Caliper	0.40	0.25	2.35	2.25	4.80	4.90
	5 Nearest Neighbors in Caliper	0.45	0.25	2.20	1.90	4.60	4.55
		Propensity Score Matching (7 Variables)					
Propensity Score Distance	1 Nearest Neighbor in Caliper	0.20	0.40	1.80	2.35	3.90	5.35
	2 Nearest Neighbors in Caliper	0.65	0.15	2.65	2.30	5.05	5.45
	3 Nearest Neighbors in Caliper	0.70	0.25	3.10	2.25	5.75	5.90
	4 Nearest Neighbors in Caliper	1.10	0.15	3.30	1.75	6.00	4.65
	5 Nearest Neighbors in Caliper	1.00	0.15	3.50	1.45	6.50	4.00
	All in Caliper	1.40	0.15	4.20	1.00	7.45	3.25
Mahalanobis Distance	1 Nearest Neighbor in Caliper	0.25	0.55	2.00	2.75	4.05	6.25
	2 Nearest Neighbors in Caliper	0.50	0.20	1.80	2.20	4.20	5.65

Continued on Next Page...

Table 1.6 – Continued

		Theoretical CDF (%)					
		0.5	99.5	2.5	97.5	5	95
	3 Nearest Neighbors in Caliper	0.50	0.20	2.20	2.00	3.95	5.15
	4 Nearest Neighbors in Caliper	0.55	0.25	2.45	2.20	4.30	4.95
	5 Nearest Neighbors in Caliper	0.65	0.25	2.10	2.00	4.15	4.75
		Conventional Methods					
Control Firm	LBT Control Firm	0.00	1.95	0.25	7.45	1.10	12.60
	5 Random Control Firms	0.15	1.15	0.90	4.90	2.15	10.10
	Ibbotson's RATS	0.00	13.10	0.00	31.40	0.00	43.80
Calendar Time Portfolios	Equal Weghted w/ FF 4-factor	0.00	17.20	0.00	36.95	0.25	49.70
	Value Weighted w/ FF 4-factor	0.25	2.90	1.50	8.60	2.95	14.70

Table 1.7: Specification Tests - Time-Clustered Samples

We select 200 firms on a randomly picked year from our sample and assign each of these an event. For each event, we assign an event month of January. This produces time clustered samples. We then calculate long-run abnormal returns using the methods tabulated below. We repeat this exercise 2,000 times and compute the percentage of t-statistics from each method that falls in the 5%, 2.5%, or 0.5% tails on the left and right sides of the corresponding t-distribution. For propensity score matching, we use a caliper of 0.01 standard deviations of the predicted values from the corresponding logistic regression to select our matches from. If this doesn't produce any matches, we use an additional caliper of 0.02 standard deviations from which one match is selected. The variables used to calculate estimated propensity scores are Size, Book-Market, Momentum, Leverage, and Return on Assets, and their squares.

		Theoretical CDF (%)					
		0.5	99.5	2.5	97.5	5	95
12-Month Returns							
		Propensity Score Matching (5 Variables)					
Propensity Score Distance	1 Nearest Neighbor in Caliper	0.75	0.80	3.75	2.45	6.40	5.55
	2 Nearest Neighbors in Caliper	1.10	0.45	4.05	2.30	7.35	4.65
	3 Nearest Neighbors in Caliper	1.35	0.25	4.10	1.75	6.85	4.20
	4 Nearest Neighbors in Caliper	1.45	0.20	4.40	1.25	7.75	3.65
	5 Nearest Neighbors in Caliper	1.35	0.25	4.95	1.55	8.05	3.60
	All in Caliper	2.05	0.05	5.30	1.45	8.70	3.00
Mahalanobis Distance	1 Nearest Neighbor in Caliper	0.60	0.25	2.80	2.30	6.60	4.75
	2 Nearest Neighbors in Caliper	1.30	0.10	4.85	1.60	8.15	3.85

Continued on Next Page...

Table 1.7 – Continued

		Theoretical CDF (%)					
		0.5	99.5	2.5	97.5	5	95
	3 Nearest Neighbors in Caliper	1.55	0.05	5.90	1.85	9.50	3.85
	4 Nearest Neighbors in Caliper	1.75	0.10	6.30	2.00	10.60	3.95
	5 Nearest Neighbors in Caliper	2.00	0.10	6.85	1.80	11.20	4.05
		Conventional Methods					
Control Firm	LBT Control Firm	0.75	0.40	3.25	2.00	6.10	4.60
	5 Random Control Firms	1.65	0.30	4.05	1.20	7.05	3.15
	Ibbotson's RATS	8.75	49.80	15.40	57.90	18.85	60.75
Calendar Time Portfolios	Equal Weghted w/ FF 4-factor	0.45	1.00	2.00	6.05	3.80	10.20
	Value Weighted w/ FF 4-factor	0.50	0.45	2.20	2.30	4.75	5.20
36-Month Returns							
		Propensity Score Matching (5 Variables)					
Propensity Score Distance	1 Nearest Neighbor in Caliper	0.35	0.30	2.20	2.35	4.70	4.65
	2 Nearest Neighbors in Caliper	0.75	0.30	3.35	1.95	6.20	4.70
	3 Nearest Neighbors in Caliper	0.80	0.20	3.35	1.95	6.80	3.95

Continued on Next Page...

Table 1.7 – Continued

		Theoretical CDF (%)					
		0.5	99.5	2.5	97.5	5	95
	4 Nearest Neighbors in Caliper	1.00	0.15	4.30	1.95	7.50	4.00
	5 Nearest Neighbors in Caliper	1.35	0.15	4.95	1.60	8.15	3.70
	All in Caliper	2.40	0.00	6.75	1.00	10.15	2.90
Mahalanobis Distance	1 Nearest Neighbor in Caliper	0.30	0.35	2.15	2.25	5.25	5.35
	2 Nearest Neighbors in Caliper	0.70	0.30	3.50	1.55	6.00	3.95
	3 Nearest Neighbors in Caliper	1.00	0.10	3.80	1.85	7.40	3.75
	4 Nearest Neighbors in Caliper	1.25	0.00	4.60	1.45	7.55	3.80
	5 Nearest Neighbors in Caliper	1.45	0.00	5.05	1.50	8.00	3.85
		Conventional Methods					
Control Firm	LBT Control Firm	0.25	0.50	1.95	2.70	4.65	5.30
	5 Random Control Firms	1.45	0.15	3.90	1.60	6.45	4.20
	Ibbotson's RATS	0.20	74.75	0.40	79.95	0.85	82.30
Calendar Time Portfolios	Equal Weghted w/ FF 4-factor	0.10	8.05	0.40	24.05	1.50	34.05
	Value Weighted w/ FF 4-factor	0.10	0.55	1.20	2.55	2.65	5.35

Continued on Next Page…

Table 1.7 – Continued

		Theoretical CDF (%)					
		0.5	99.5	2.5	97.5	5	95
60-Month Returns							
		Propensity Score Matching (5 Variables)					
Propensity Score Distance	1 Nearest Neighbor in Caliper	0.25	0.20	2.10	2.25	5.20	5.10
	2 Nearest Neighbors in Caliper	0.35	0.20	2.50	2.00	5.65	4.80
	3 Nearest Neighbors in Caliper	0.65	0.10	3.00	1.70	6.50	4.45
	4 Nearest Neighbors in Caliper	1.10	0.05	4.00	1.60	7.30	4.10
	5 Nearest Neighbors in Caliper	1.35	0.10	4.35	1.90	8.60	4.05
	All in Caliper	2.75	0.05	7.70	0.85	11.75	2.90
Mahalanobis Distance	1 Nearest Neighbor in Caliper	0.05	0.35	2.30	2.35	4.90	4.90
	2 Nearest Neighbors in Caliper	0.45	0.15	2.60	1.70	5.55	4.40
	3 Nearest Neighbors in Caliper	0.80	0.10	3.85	1.45	7.15	3.95
	4 Nearest Neighbors in Caliper	1.20	0.00	4.45	1.25	8.00	3.60
	5 Nearest Neighbors in Caliper	1.25	0.00	4.85	1.20	8.65	3.70
		Conventional Methods					

Continued on Next Page...

Table 1.7 – Continued

		Theoretical CDF (%)					
		0.5	99.5	2.5	97.5	5	95
Control Firm	LBT Control Firm	0.10	0.45	2.25	2.35	4.75	5.45
	5 Random Control Firms	1.20	0.05	5.00	1.70	8.05	3.55
	Ibbotson's RATS	3.60	82.70	4.35	87.00	4.45	88.70
Calendar Time Portfolios	Equal Weghted w/ FF 4-factor	0.00	18.75	0.10	40.85	0.40	51.40
	Value Weighted w/ FF 4-factor	0.25	0.75	1.45	3.60	2.75	7.40

Table 1.8: Specification Tests - Serial-Correlated Samples

We randomly select 100 firm-years from our sample and assign each of these an event. For each such event, we impose another event on the firm in the following calendar year. The event month is constrained to January. We then calculate long-run abnormal returns using the methods tabulated below. We repeat this exercise 2,000 times and compute the percentage of t-statistics from each method that falls in the 5%, 2.5%, or 0.5% tails on the left and right sides of the corresponding t-distribution. For propensity score matching, we use a caliper of 0.01 standard deviations of the predicted values from the corresponding logistic regression to select our matches from. If this doesn't produce any matches, we use an additional caliper of 0.02 standard deviations from which one match is selected. The variables used to calculate estimated propensity scores are Size, Book-Market, Momentum, Leverage, and Return on Assets, and their squares.

		Theoretical CDF (%)					
		0.5	99.5	2.5	97.5	5	95
12-Month Returns							
		Propensity Score Matching (5 Variables)					
Propensity Score Distance	1 Nearest Neighbor in Caliper	0.40	0.55	2.30	2.80	4.90	5.05
	2 Nearest Neighbors in Caliper	0.60	0.25	3.00	2.00	5.65	4.10
	3 Nearest Neighbors in Caliper	1.00	0.10	3.70	1.45	6.85	3.40
	4 Nearest Neighbors in Caliper	0.90	0.05	4.10	1.35	7.50	3.25
	5 Nearest Neighbors in Caliper	1.05	0.05	4.50	1.05	7.45	3.40
	All in Caliper	1.65	0.00	5.25	0.95	9.05	3.00
Mahalanobis Distance	1 Nearest Neighbor in Caliper	0.25	0.50	1.85	2.65	4.25	5.35
	2 Nearest Neighbors in Caliper	0.60	0.20	3.15	2.95	5.40	6.05

Continued on Next Page...

Table 1.8 – Continued

		0.5	99.5	2.5	97.5	5	95
		Theoretical CDF (%)					
	3 Nearest Neighbors in Caliper	0.60	0.10	3.55	2.05	6.00	4.75
	4 Nearest Neighbors in Caliper	0.75	0.00	4.00	1.75	6.35	4.55
	5 Nearest Neighbors in Caliper	0.95	0.05	3.55	1.65	6.65	4.60
		Conventional Methods					
Control Firm	LBT Control Firm	0.60	0.50	3.00	2.35	4.85	5.55
	5 Random Control Firms	0.90	0.00	4.50	1.25	8.55	3.20
	Ibbotson's RATS	0.00	0.90	0.25	6.00	0.85	11.50
Calendar Time Portfolios	Equal Weghted w/ FF 4-factor	0.00	4.60	0.15	14.00	0.25	24.85
	Value Weighted w/ FF 4-factor	0.35	0.55	2.15	1.80	4.05	4.35
36-Month Returns							
		Propensity Score Matching (5 Variables)					
Propensity Score Distance	1 Nearest Neighbor in Caliper	0.55	0.75	3.85	2.85	7.65	5.60
	2 Nearest Neighbors in Caliper	1.25	0.60	4.40	2.90	8.65	6.25
	3 Nearest Neighbors in Caliper	2.20	0.40	6.25	2.65	10.10	4.90

Continued on Next Page...

Table 1.8 – Continued

		Theoretical CDF (%)					
		0.5	99.5	2.5	97.5	5	95
	4 Nearest Neighbors in Caliper	2.70	0.40	7.05	2.10	11.35	5.05
	5 Nearest Neighbors in Caliper	2.90	0.35	8.55	2.10	12.25	4.55
	All in Caliper	5.70	0.30	10.95	1.55	15.45	3.65
Mahalanobis Distance	1 Nearest Neighbor in Caliper	0.60	0.50	3.60	3.05	6.95	6.25
	2 Nearest Neighbors in Caliper	1.00	0.65	3.90	2.75	7.65	6.50
	3 Nearest Neighbors in Caliper	1.60	0.50	5.40	3.00	9.55	6.25
	4 Nearest Neighbors in Caliper	2.00	0.40	6.40	2.50	10.35	5.65
	5 Nearest Neighbors in Caliper	2.55	0.40	6.75	2.00	10.65	5.70
		Conventional Methods					
Control Firm	LBT Control Firm	0.95	0.60	3.75	2.80	6.80	6.30
	5 Random Control Firms	2.30	0.50	6.95	1.85	11.35	4.35
	Ibbotson's RATS	0.05	8.60	0.55	20.50	1.10	29.85
Calendar Time Portfolios	Equal Weghted w/ FF 4-factor	0.00	7.95	0.00	22.25	0.15	34.70
	Value Weighted w/ FF 4-factor	0.40	0.75	2.25	3.05	4.70	5.30

Continued on Next Page…

Table 1.8 – Continued

		Theoretical CDF (%)					
		0.5	99.5	2.5	97.5	5	95
		60-Month Returns					
		Propensity Score Matching (5 Variables)					
Propensity Score Distance	1 Nearest Neighbor in Caliper	0.70	0.80	3.30	3.05	6.10	5.80
	2 Nearest Neighbors in Caliper	1.70	0.55	5.45	3.20	8.40	5.90
	3 Nearest Neighbors in Caliper	2.30	0.40	6.40	2.30	9.95	6.25
	4 Nearest Neighbors in Caliper	2.45	0.35	7.75	2.55	11.80	5.25
	5 Nearest Neighbors in Caliper	3.05	0.25	8.15	2.10	12.15	5.00
	All in Caliper	6.75	0.20	13.40	1.60	17.70	3.35
Mahalanobis Distance	1 Nearest Neighbor in Caliper	0.65	0.50	2.95	2.70	6.00	6.25
	2 Nearest Neighbors in Caliper	0.80	0.45	4.30	3.15	7.35	6.85
	3 Nearest Neighbors in Caliper	1.60	0.35	5.35	2.85	9.15	5.90
	4 Nearest Neighbors in Caliper	2.20	0.30	6.55	2.70	9.95	5.95
	5 Nearest Neighbors in Caliper	2.70	0.40	7.05	2.55	11.40	5.50
		Conventional Methods					

Continued on Next Page...

Table 1.8 – Continued

		Theoretical CDF (%)					
		0.5	99.5	2.5	97.5	5	95
Control Firm	LBT Control Firm	0.85	0.90	2.75	3.25	6.65	6.65
	5 Random Control Firms	2.55	0.50	7.70	2.25	11.85	4.90
	Ibbotson's RATS	0.05	13.70	0.15	28.65	0.65	38.95
Calendar Time Portfolios	Equal Weghted w/ FF 4-factor	0.00	11.20	0.05	27.25	0.20	38.45
	Value Weighted w/ FF 4-factor	0.45	1.05	2.25	3.85	4.45	7.20

Figure 1.1: Simulated Biased Distribution of Observed Covariate

This figure shows the empirical distribution of the observable X for 10,000 units in simulated data, with a quarter of the units randomly assigned to the treatment group. X for each control unit is drawn from an $N(0,1)$ distribution, while X for each treated unit is drawn from an $N(1,1)$ distribution.

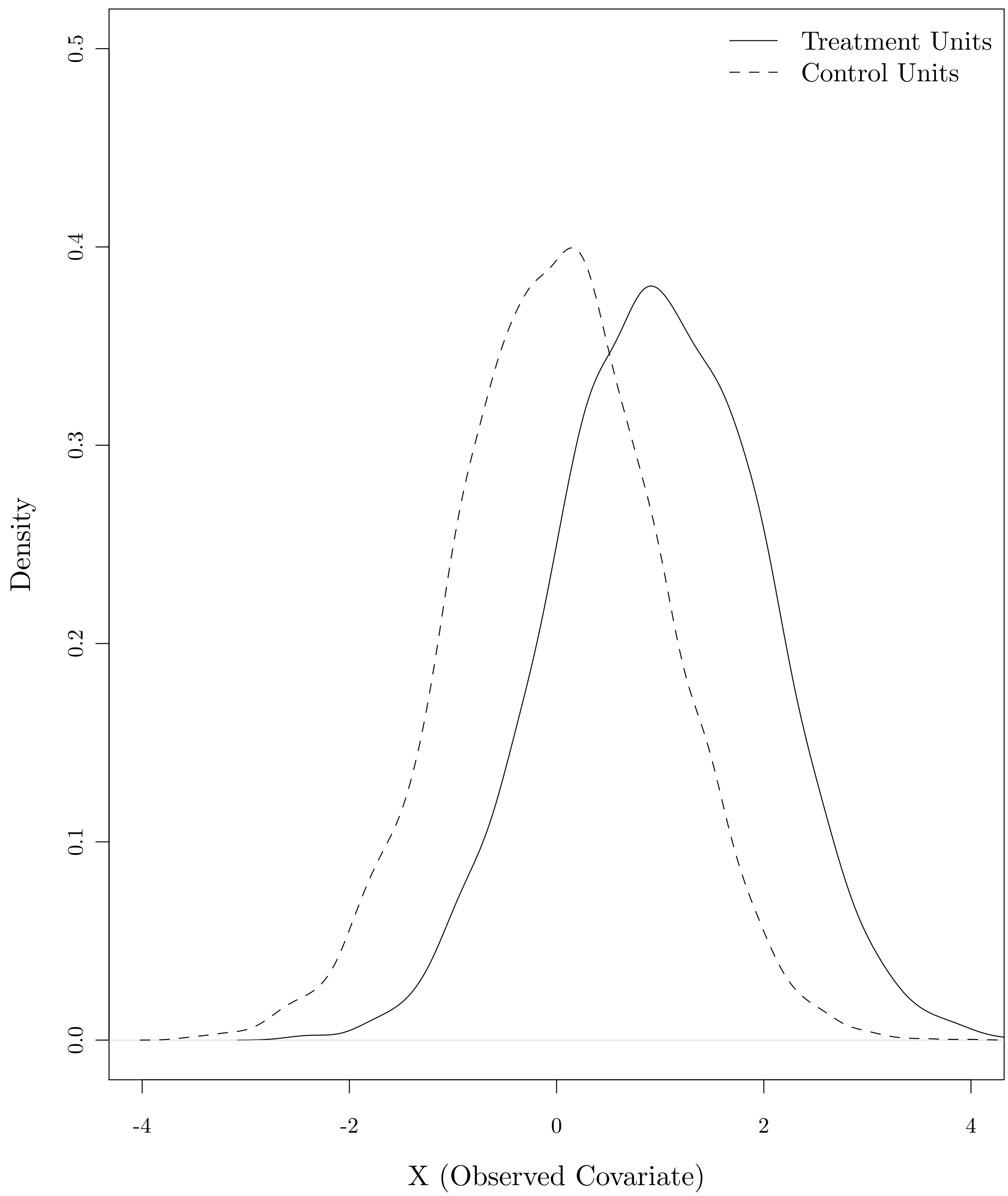


Figure 1.2: Simulated Biased Distribution of Outcome

This figure shows the empirical distribution of the outcome Y for 10,000 units in simulated data, with a quarter of the units randomly assigned to the treatment group. X for each control unit is drawn from an $N(0,1)$ distribution, while X for each treated unit is drawn from an $N(1,1)$ distribution. Y is generated according to $Y_i = T_i + X_i + \epsilon_i$, where T_i denotes treatment status and ϵ_i is drawn from $N(0, 0.05)$.

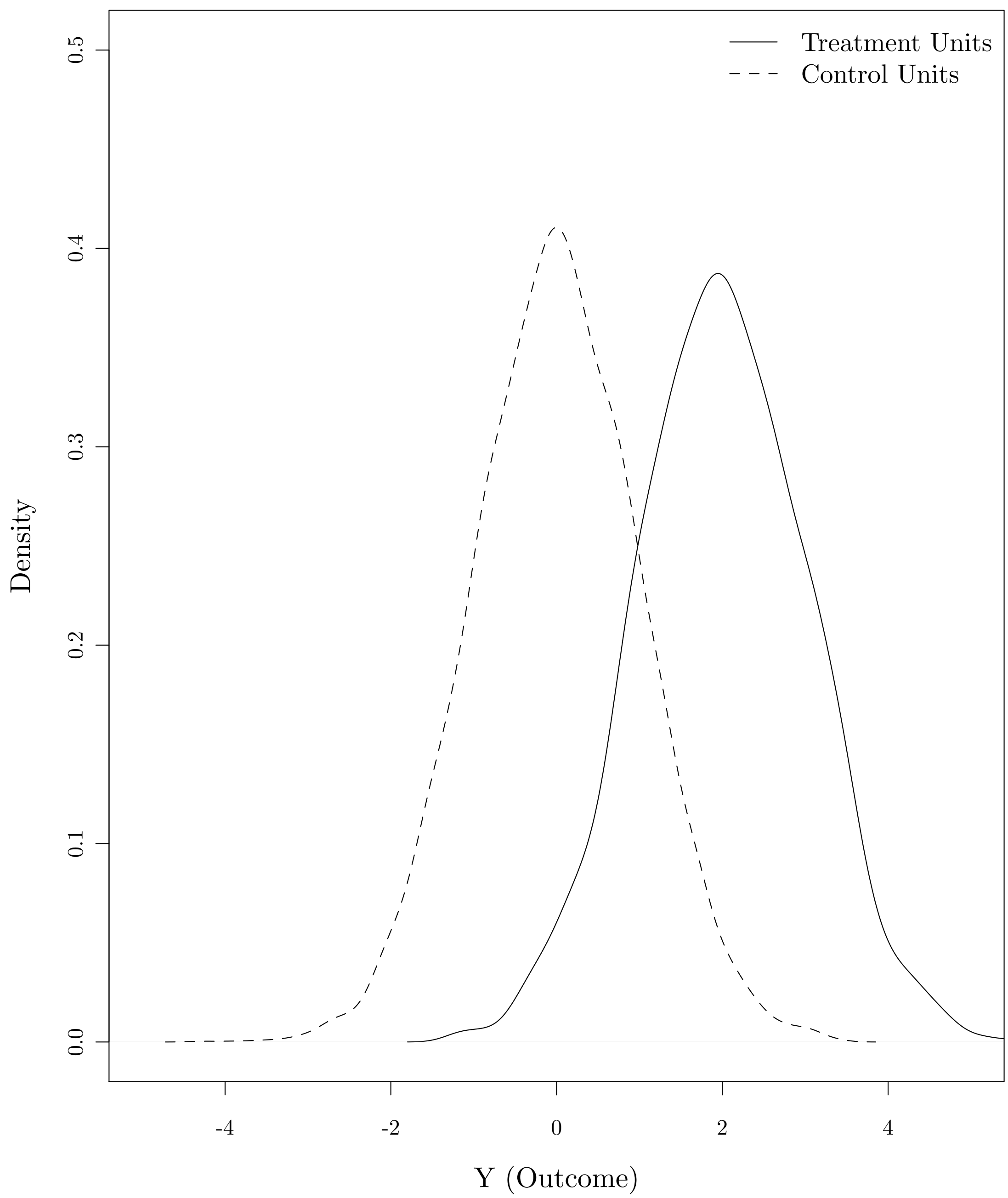


Figure 1.3: Outcome for Simulated Dataset

This figure shows the outcome Y for 10,000 units in simulated data as a function of the observed covariate X, with a quarter of the units randomly assigned to the treatment group. X for each control unit is drawn from an $N(0,1)$ distribution, while X for each treated unit is drawn from an $N(1,1)$ distribution. Y is generated according to $Y_i = T_i + X_i + \epsilon_i$, where T_i denotes treatment status and ϵ_i is drawn from $N(0, 0.05)$.

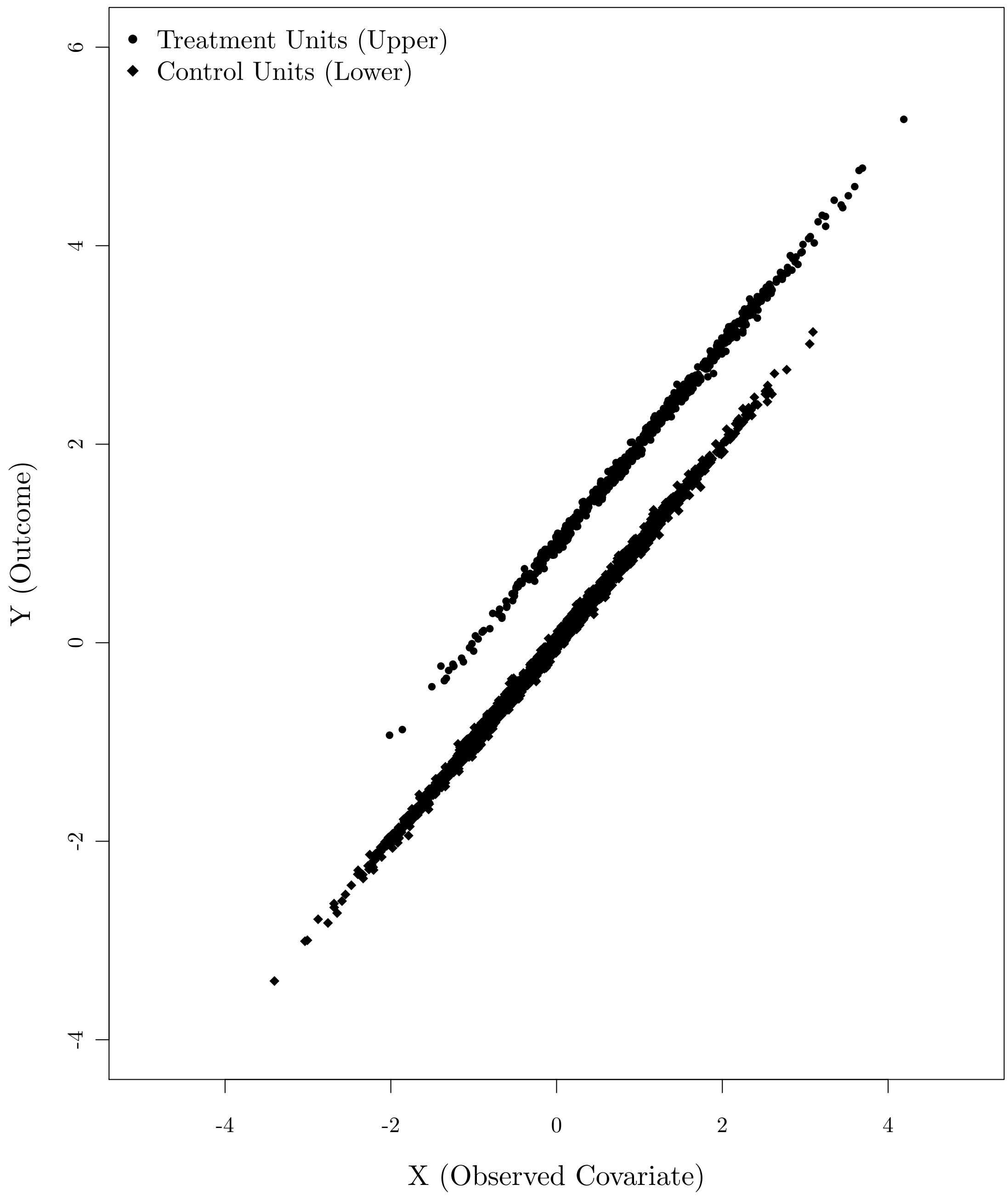


Figure 1.4: Number of Matches within Caliper

This figure shows the number of matches that are obtained through propensity score matching on random samples within a specified caliper. We use an initial caliper of 0.01 standard deviations of the estimated propensity scores (the predicted values of the logistic regression). If no match is found, we use a secondary caliper of 0.02 standard deviations and select the closest match. The horizontal axis shows the number of matches. The vertical column shows the relative frequency as a percentage, with the final bar showing the frequency for >100 matches. The total number of observations are 400,000 of which 1% have no matches (not shown here).

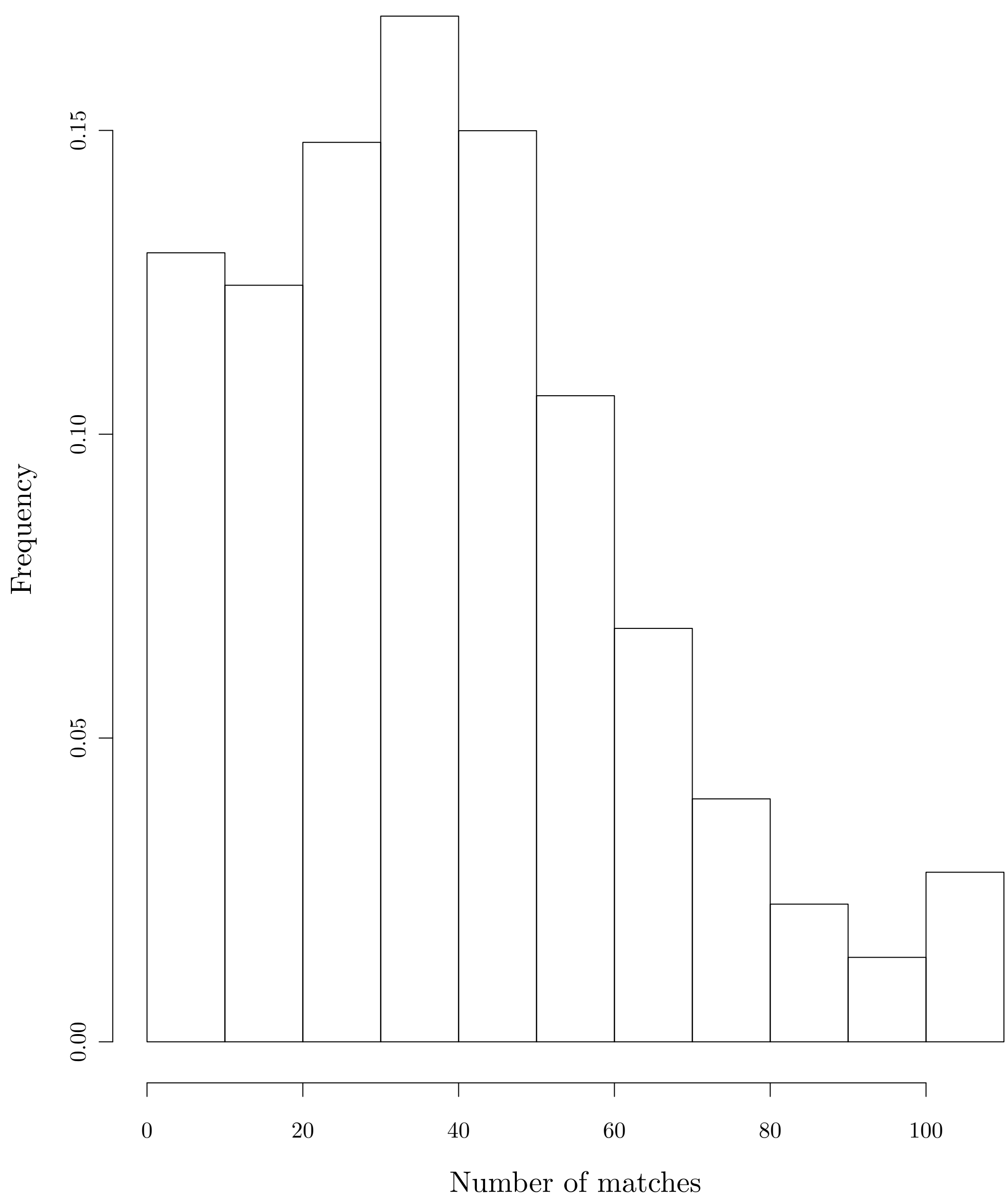


Figure 1.5: Empirical Distribution of 12-month Buy-and-Hold Returns

This figure shows the empirical distribution of 12-month buy-and-hold returns for single firms and 5-firm portfolios. The data are generated using the CRSP monthly database from 1984-2004. For single firms, 10,000 observation points (firm-months) are randomly chosen and the 12-month buy-and-hold returns are calculated for the corresponding firm starting from that month. For 5-firm portfolios, the above process is repeated 5 times and the arithmetic mean is calculated over the 5 iterations generating 10,000 values of average 12-month buy-and-hold returns. The horizontal axis denotes the return (-100% - 500%) while the vertical axis shows the estimated density using kernel density estimation. The distribution extends a little to the left of -1 due to the estimation process; the actual data stops at -1.

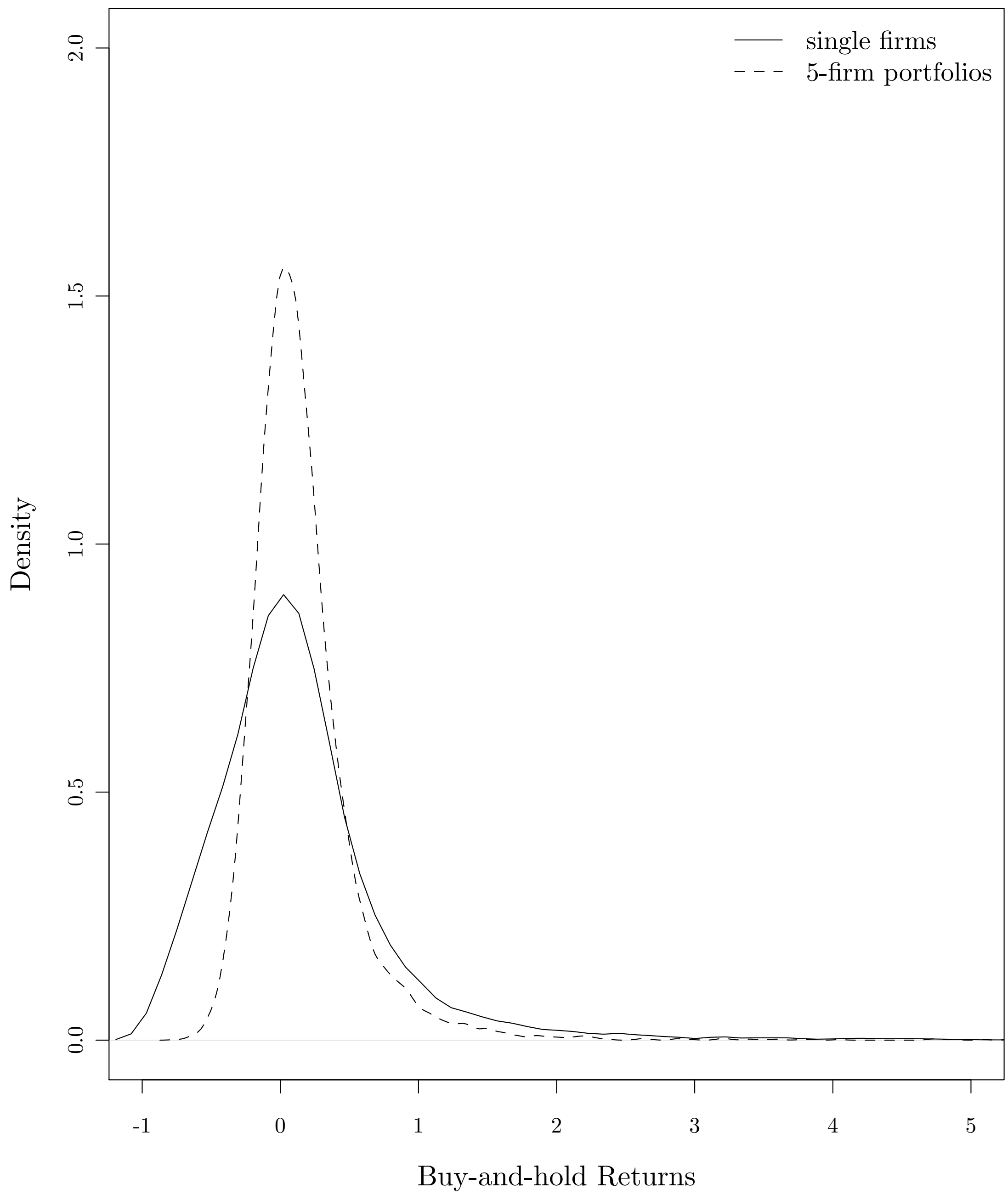


Figure 1.6: Empirical Distribution of 36-month Buy-and-Hold Returns

This figure shows the empirical distribution of 36-month buy-and-hold returns for single firms and 5-firm portfolios. The data are generated using the CRSP monthly database from 1984-2004. For single firms, 10,000 observation points (firm-months) are randomly chosen and the 36-month buy-and-hold returns are calculated for the corresponding firm starting from that month. For 5-firm portfolios, the above process is repeated 5 times and the arithmetic mean is calculated over the 5 iterations generating 10,000 values of average 36-month buy-and-hold returns. The horizontal axis denotes the return (-100% - 500%) while the vertical axis shows the estimated density using kernel density estimation. The distribution extends a little to the left of -1 due to the estimation process; the actual data stops at -1.

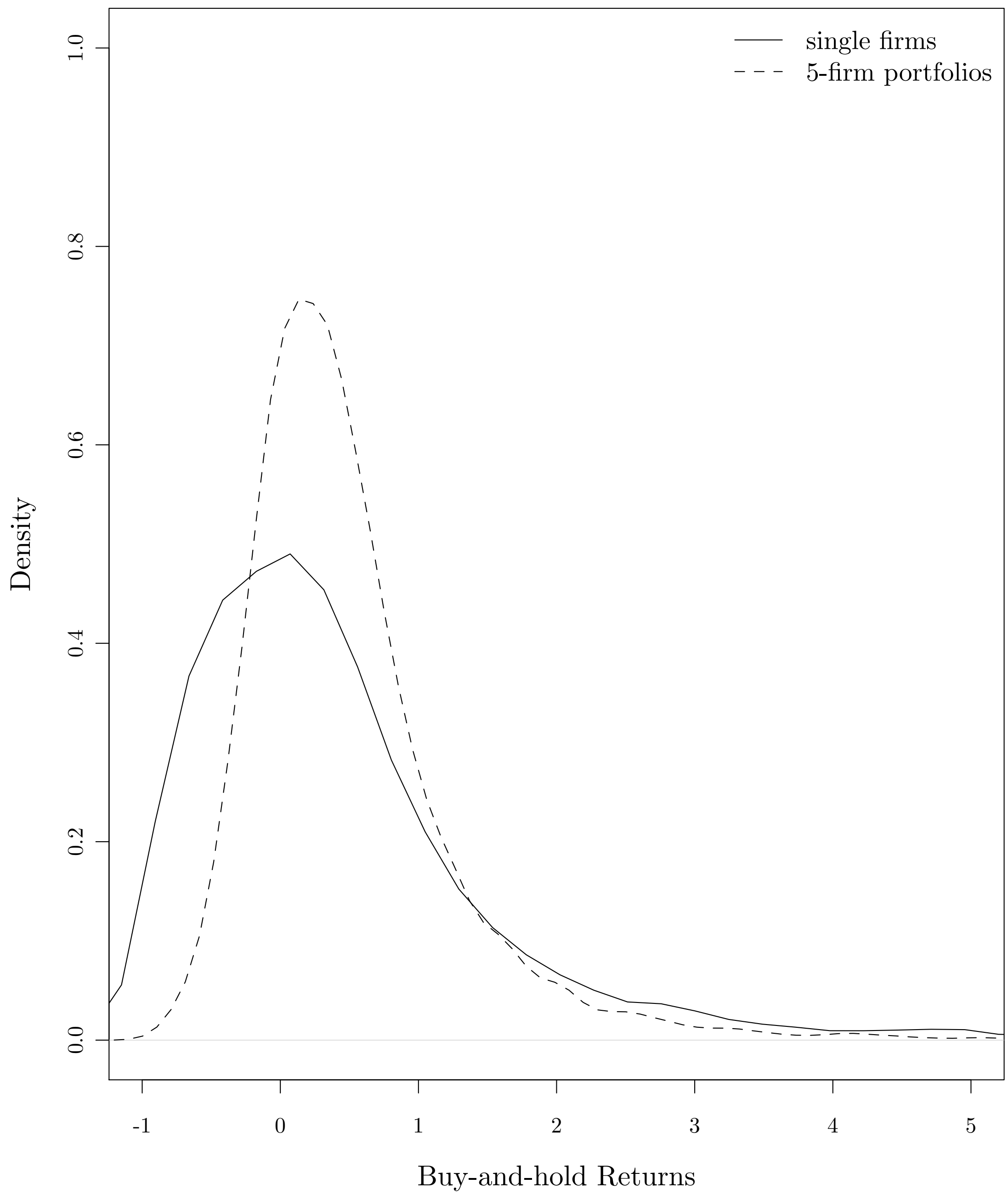


Figure 1.7: Empirical Distribution of 60-month Buy-and-Hold Returns

This figure shows the empirical distribution of 60-month buy-and-hold returns for single firms and 5-firm portfolios. The data are generated using the CRSP monthly database from 1984-2004. For single firms, 10,000 observation points (firm-months) are randomly chosen and the 60-month buy-and-hold returns are calculated for the corresponding firm starting from that month. For 5-firm portfolios, the above process is repeated 5 times and the arithmetic mean is calculated over the 5 iterations generating 10,000 values of average 60-month buy-and-hold returns. The horizontal axis denotes the return (-100% - 500%) while the vertical axis shows the estimated density using kernel density estimation. The distribution extends a little to the left of -1 due to the estimation process; the actual data stops at -1.

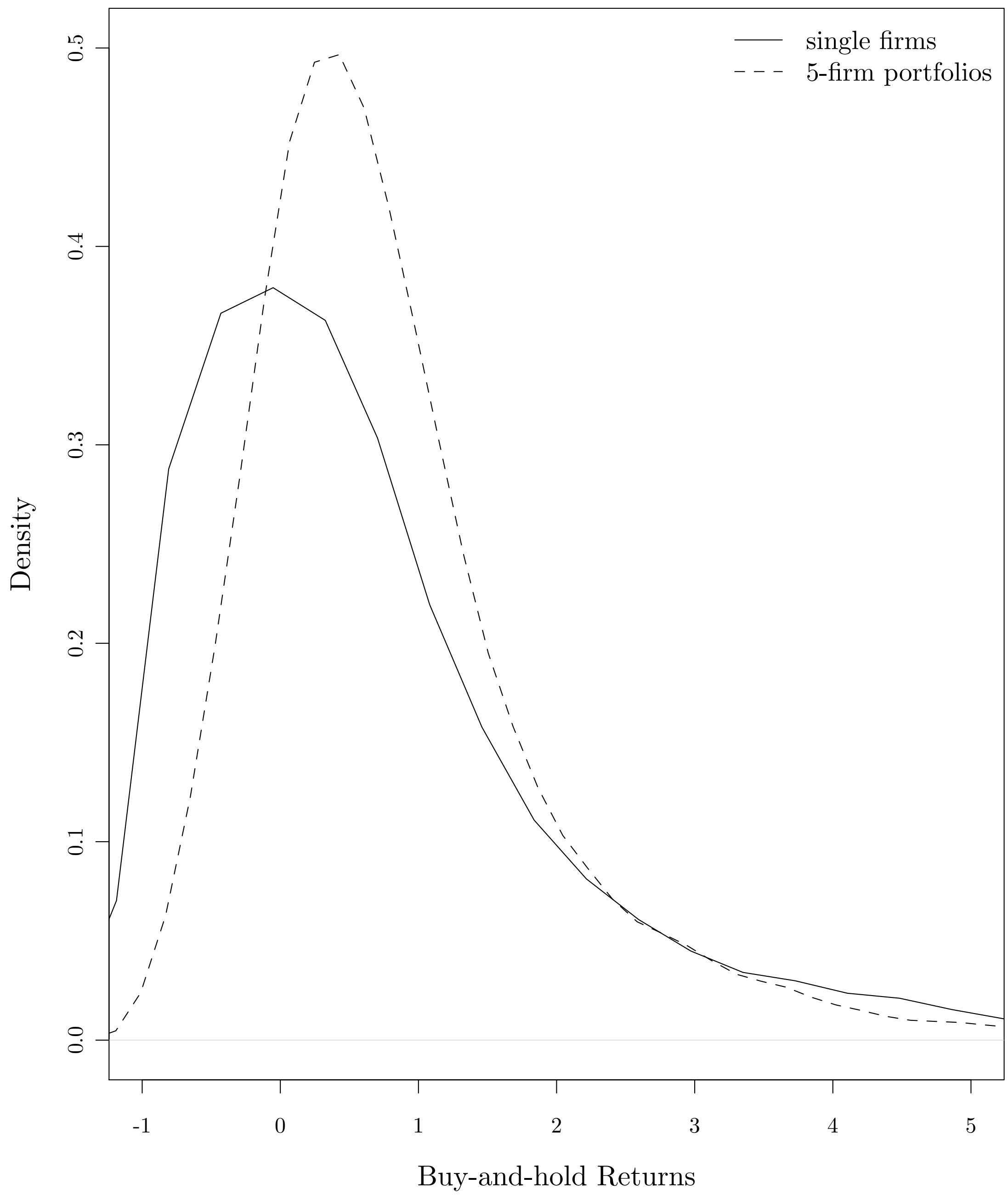
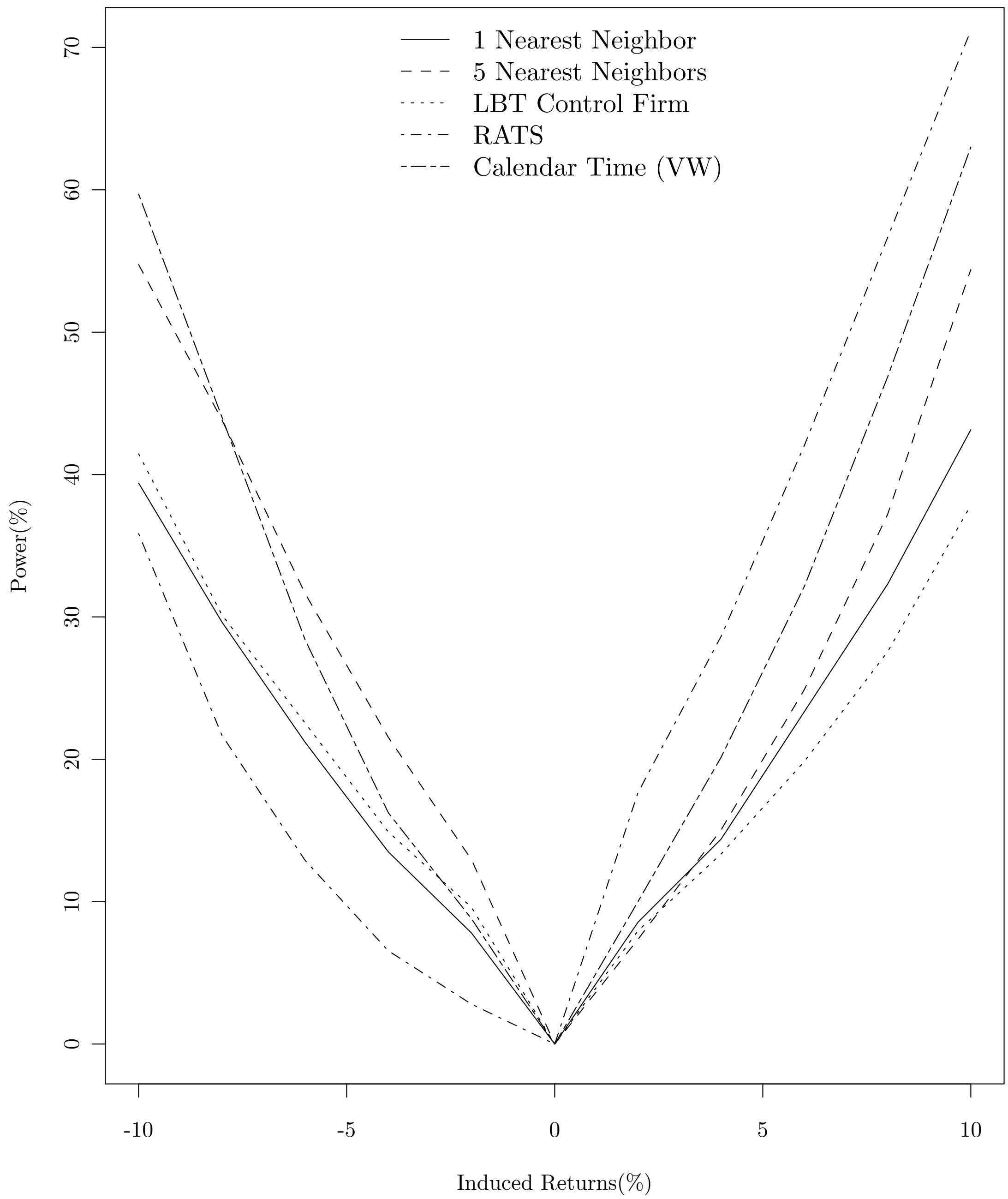


Figure 1.8: Power in Random Samples

This figure shows the power of various techniques in detecting induced abnormal returns in random samples. For each iteration in Table 1.3 we increase(decrease) the 12-month buy-and-hold return by a fixed amount for each event. We then check for the fraction of t-statistics in each method that fall in the right(left) 5% tail of the corresponding t-distribution; i.e. we check for power at the 5% significance level using a 1-tailed t-test. The induced returns are from 0 to +(-)10%. For propensity score matching, we use the Mahalanobis metric.



Chapter 2

Long-Run Abnormal Returns following Open-Market Share Repurchases

2.1 Introduction

According to Securities Data Corporation's Mergers & Acquisition database, the last five years have seen U.S. corporations announce share repurchase programs at estimated values of more than \$300 bn on an average annually, at least 90% of which were open-market. There are contending theories for the motivation behind the increasing use of share repurchase programs. One suggests that managers use changes in payout policy to convey information about future performance (Miller & Modigliani (1961), Miller & Rock (1985)). However, as documented by Gong et al. (2008), evidence pointing in this direction could be a result of earnings management in the quarter prior to a repurchase announcement, because of which subsequent quarters appear particularly good. A competing theory states that firms, in a bid to avoid value-destroying activities, distribute cash flow in excess of their investment opportunities via changes in payout policy that may involve repurchase programs (Jensen

(1986)). Grullon and Michaely (2002) find evidence in support of this theory in the form of lowered systematic risk following repurchases. In a survey of 384 financial executives, Brav et al. (2005) attempt to unearth the true underlying motivation through in-depth interviews asking managers a range of questions about their payout policies. In this survey, about 85% of the managers expressed the belief that there are negative consequences to reducing dividends and that it is essential to maintain consistency with prior policy regarding dividends. The corresponding number for repurchases as a payout tool was around 20%. Regardless of the underlying cause, repurchases are thus seen as a more flexible alternative to dividends when decisions about payouts are made. More importantly, Brav et al. (2005) also report that almost 90% of the managers surveyed strongly agreed with announcing repurchases when the market price of the stock is below its perceived true value. In addition, there are several studies that confirm managers' ability to time the market with repurchase announcements. This is most often done by examining returns of stocks for up to five years after the announcement. Ikenberry et al. (1995) examine long-run performance of stocks following open-market share repurchases for the period 1980-1990 and find that the post-announcement average abnormal buy-and-hold return over the next four years is 12.1%. Using data on open market programs announced by firms listed on the Toronto Stock Exchange from 1989 to 1997, Ikenberry et al. (2000) find abnormal returns of 0.59% per month over a three-year period following the announcement. McNally and Smith (2007) use a unique Canadian dataset of all repurchase programs by firms listed on the Toronto Stock Exchange from 1987 to 2000 and document 4-6% annual abnormal returns over a three-year period. However, these returns vanished upon excluding penny stocks. Chan et al. (2007), using U.S. repurchase announcements from 1980 to 1996, finds buy-and-hold abnormal returns of 24% and monthly abnormal returns of about 0.3% over a four-year period. Gong et al. (2008) find significant negative correlation between pre-announcement abnormal accruals and post-announcement long-run returns. Peyer and Vermaelen (2009) conduct their analysis for U.S. repurchases confirmed by LexisNexis from 1991 to 2001 and document 2.67%, 10.54%, 18.60%, and 24.25% cumulative abnormal returns over one, two, three, and four years respectively, and corresponding monthly abnormal returns ranging from 0.45-0.55%. Overall, there seems to be substantial evidence in support of managers' ability to time the market by identifying undervalued stocks. An important piece of evidence against this effect was documented by Gong et al. (2008) who found that downward earnings management before the repurchase announcement has a significant negative association with future stock performance.

This chapter is an attempt at a more accurate calculation of long-run abnormal returns following open-market share repurchases. We identify certain characteristics that are more common among repurchasing firms. We document that firms announcing repurchases are larger, are rarely in the top book-to-market decile and have usually had high return on assets (calculated using net income) in the prior calendar year. We form an aggregated measure using the three variables just mentioned and assign events to specific quantiles of this new variable. The results of these simulations suggest that the best approach to calculating long-run abnormal returns in this case is matching based on propensity scores, Consequently, we do find the presence of buy-and-hold abnormal returns of about 3%, 7.5%, and 12% over one-, three-, and five-year periods. These returns are much smaller compared to that documented in prior studies. In addition, practically all of the abnormal returns are concentrated amongst firms that fall in the bottom half of ROA, which could tie in with the findings of Gong et al.

The rest of this study is organized as follows. Section 2 describes the data. Section 3 discusses the methodology and the results. Concluding remarks are offered in Section 4.

2.2 Data

Our initial sample consists of all U.S. repurchases present in SDC Platinum's *Mergers & Acquisitions* database classified as 'Open'. The entire database spans 1980 to 2009, while repurchases classified as open-market begin in 1984. Since we focus on long-run returns for up to five years, we restrict our sample to Dec-2004. In addition, we require that:

- The firm must be present in CRSP with a share code of 10 or 11 (ordinary common shares) and an exchange code of 1, 2. or 3 (NYSE, AMEX, or Nasdaq) at the time of announcement, as well as have a market value as of December of the prior year.

- The firm must be present in Compustat Annual Database with non-missing Assets (AT), Common/Ordinary Equity (CEQ), Liabilities (LT), Cash & Short Term Investments (CHE), Net Income (NI), and Total Sales (SALE) as of December[1] of the prior year.

[1]In most cases we do find balance sheet data reported in December. Otherwise, we select the

These criteria give us 7,377 repurchase announcements from 1984 to 2004. Dittmar (2000) finds, using a Tobit model with several variables, that larger firms and firms with higher cash flow are more likely to repurchase stock. She also finds that firms use stock repurchases to distribute excess cash to stockholders, after controlling for investment opportunities, and that a firm's leverage relative to its target ratio has a minor but significant effect on repurchase decisions. On the other hand, Ikenberry et al. (1995, 2000) and Peyer & Vermaelen (2000) find that value stocks generate the highest long-run abnormal returns following announcement. If managers do possess market timing skills, this raises the possibility that firms with high book-to-market values may be more likely to announce repurchase programs to take advantage of potentially higher mispricing.

Table 2.1 presents summary statistics of repurchasing firms compared to non-repurchasing firms in terms of various firm characteristics. Each of these variables is normalized to have a zero-mean and unit-standard deviation for easier interpretation. We retain these variables in this form for the remainder of the study. We classify a firm as 'Repurchasing' at December of a particular year if it has at least one repurchasing announcement during the next calendar year. Otherwise, the firm is classified as 'Non-Repurchasing'.

As seen, all six variables are significantly different between the repurchasing and non-repurchasing samples. Among the six, based on the magnitude of the t-statistics, three factors appear to be more important: size, book-market, and ROA. The t-statistic for the difference in means of book-market ratio between the two samples

closest available month in the same calendar year.

is -26.5 and that for ROA is 53.4. The value for size isn't as high at 12, but that is likely to be caused by the extreme skewness of the size variable. To get a better understanding of the distribution of the variables across the two samples, we look at Table 2.2. This table presents the fraction of repurchasing firms (in %) in each decile of the individual firm characteristics. The decile cut-off points are estimated separately for each year. The top row shows that 2.2% of the firms in the repurchasing sample lie in the lowest size decile. This number steadily increases to 20.5% of the repurchasing firms in the highest size decile. Clearly, larger firms are more likely to issue a repurchase announcement. The second row shows that firms with very high or very low book-market ratios are unlikely to announce a repurchase program, since the lowest and the highest book-market deciles contain only 7.1% and 4.6% of the firms respectively. The top two ROA deciles contain almost 30% of the repurchasing firms, while the bottom two deciles contain only about 5%, implying that firms with high ROA are more likely to announce repurchases. There isn't as distinctive a pattern for the remaining variables as for the three just described. For leverage, the decile with the highest concentration of firms has 12%, while the decile with the lowest concentration of firms has 8%, making the difference 4%. The corresponding difference for the cash variable is 4.4%, and that for tat is 3.3%.

We interpret these observations to mean that repurchasing firms are more likely to be large and highly profitable, and they are less likely to have extreme book-market values, especially on the higher side. Using this interpretation, we construct a new variable V_{net} as $size + roa$ - $book - market$. The distribution of repurchasing firms across deciles of V_{net} is shown in the last row of Table 2.2. Clearly, as V_{net} for a firm gets higher, it is more likely to announce a repurchase.

2.3 Methodology & Results

2.3.1 Sample Characteristics

Based on the univariate analysis in the preceding section, we found that firms with high values of size and ROA are more likely and firms with high values of book-market are less likely to announce repurchase programs. Other variables, namely leverage, cash, and TAT also influenced likelihood of participation in a repurchase, but not to the same extent. We now confirm these observations using the following logistic regression:

$$
\begin{aligned}
Tr_{it} =& \alpha + \beta_1 size_{it} + \beta_1' size_{it}^2 + \beta_2 book - market_{it} + \beta_2' book - market_{it}^2 + \\
& \beta_3 leverage_{it} + \beta_3' leverage_{it}^2 + \beta_4 roa_{it} + \beta_4' roa_{it}^2 + \\
& \beta_5 cash_{it} + \beta_5' cash_{it}^2 + \beta_6 tat_{it} + \beta_6' tat_{it}^2 + \epsilon_{it},
\end{aligned}
$$

where $Tr_{it} = 1$ denotes participation in a repurchase announcement by firm i in year t. This is shown in Table 2.3. We present results from a completely linear model as well for comparison. Since the variables are all normalized to the same scale, the estimates can be used to compare the relative importance of each in the estimation of the predicted probability.[1] We see that allowing for non-linearities in the model significantly alters the estimates of some of the coefficients. The estimate for *size* increases by 4 times, the estimate for *book_ market* increases by about 25% and the estimate of *ROA* increases by about 25% as well. The estimate for *cash* changes sign,

[1] The precise distributions are, of course, different. Any interpretations drawn based on these estimates are therefore only approximate.

while the other two coefficient estimates undergo smaller changes. Furthermore, the squared terms are all highly significant. We proceed with the non-linear model for the rest of the analysis. We should note, however, that since for most of the observations, the values of the independent variables lie between -1 and 1, the effect of the squared terms is only felt on a small portion of the population. This is clearly evident for the *size* variable for which the maximum value in the population is 68, but about 98% of the size values lie between -0.15 and 1. Since the maximum value for repurchasing firms is 32, the regression model tries to reduce the sum of squared errors by lowering the estimate of the coefficient, as seen in the estimate of *size* for the linear model. Thus, having the squared term can significantly improve the estimate of the linear term, particularly by picking up extreme values in either direction (or both, as in the case of $book - market$).

In this model, the estimates for the coefficients of *size*, *book-market*, and *ROA* are much larger in magnitude than the rest of the variables. To obtain a probabilistic interpretation of these estimates, we plot the predicted values probability obtained using the regression estimates on the mean values of the corresponding variables within deciles of the three independent variables mentioned above. This is shown in Figure 2.1. The graph for *size* shows that in the lowest size decile, the average firm, with values of independent variables determined by the mean values of the variables for all firm-years within that size decile,[2] has a probability of 3.2% of participating in a repurchase program. On the other hand, the average firm in the highest size decile has a probability of 12.8%. Similarly, the average firm in the lowest *ROA* decile has

[2]The values for the squared terms are obtained by squaring the mean values.

a probability of participation of 0.5%, which gets inflated to 13.3% in the highest decile. For an average firm in the second or third book_ market decile, the predicted probability is close to 8%, while in the highest decile this gets reduced to 3%. We therefore obtain a similar picture to what was earlier seen in the univariate analysis. We also see that when the deciles are divided according to the constructed variable V_{net}, the predicted probability for an average firm goes from 0.6% in the lowest decile to 15.5% in the highest decile. There are 9,514 observations in the highest V_{net} decile. Of these, 1,424 observations were associated with a repurchase announcement. This amounts to 15%, which is very close to the value obtained from the logistic regression estimate for an averge firm in the top V_{net} decile as seen in the plot. This can be taken as evidence in favor of the regression model and the predicted values are assumed to be good estimates of propensity scores.

2.3.2 Cluster Tests

We saw in Table 2.2 that almost 20% of the repurchase sample comes from the top V_{net} decile, i.e. firms with high values of *size* and *ROA* and low values of *book-market*. Firms with such characteristics have a 15.5% probability of announcing a repurchase, which is much higher compared to the average firm's 6.8% probability. This shows that repurchasing firms form non-random clusters based on these characteristics. In Section 1.5.2 on page 34, we tested on two kinds of arbitrarily selected clusters for the performance of several methods of calculating long-run abnormal returns. Since we are able, in case of repurchasing firms, to identify a clustering scheme, we perform similar tests on small samples randomly chosen from the same cluster. Similar to what we did earlier, we randomly pick 200 firm-year observations from the highest V_{net} decile and assign them an event. We then calculate t-statistics using various methods of calculating abnormal returns over 12-, 36-, and 60-months. This procedure is repeated 2,000 times. We use the predicted values from the logistic regression of Table 2.3 as the estimated propensity scores. To calculate the Mahalanobis distance, we use all 6 firm characteristics and the procedure explained on page 24.

Results of the specification tests are presented in Table 2.4. We see that almost all the propensity score matching methods are well specified and quite close to the theoretical significance limits. An exception is the *All in Caliper* rule, which owing to the large number of nearest neighbors is negatively skewed. Importantly, with one and two neighbors, we get good results with both the propensity score distance and the Mahalanobis metric. LBT's control firm, based on *size* and *book-market*, provides t-statistics with a positive skewness. 8.9% of the t-statistics fall in the right 5% tail, while 5.2% fall in the right 2.5% tail for 12-month returns. T-statistics for 36-month returns are similarly skewed as well. There is also an under-representation of values in the left tail, with only about 2% of the t-statistics falling below the 5^{th} percentile. The RATS method and the calendar time method with both equal and value weighted portfolios yield t-statistics with severe positive skewness – 9.4%, 15.1%, and 23.8% of the t-statistics from value-weighted calendar time portfolios fall in the right 5% tail for 12-, 36-, and 60-month returns respectively. Therefore, all the conventional methods tested here have a positive bias when estimating long-run abnormal returns on samples derived from firms clustered on high values of V_{net}. Since 20% of the repurchase sample are composed of such firms, there is a strong possibility for each of these methods to give positively biased estimates of long-run abnormal returns following repurchase announcements.

2.3.3 Repurchase Sample

In this subsection, we apply the methods shown and discussed in the prior sections to estimate long-run abnormal returns of repurchasing firms.

2.3.3.1 Fama-French Regression Methods

Table 2.5 shows the results of using regression-based methods with Fama-French factors. We use the full sample for Panel A and find very high and statistically significant abnormal returns using each of the methods. To reduce the effect of very cheap stocks, we discard the observations from our sample where the stock price at the month-end prior to announcement was less than or equal to \$5. These results are in Panel B. The values of the abnormal returns decrease slightly but the strength of the results doesn't change by much. We get 5.8%, 15.3%, and 21% cumulative abnormal returns over 12-, 36-, and 60-months respectively using RATS. In earlier tests, we saw that the t-statistics obtained using RATS had a severe positive skewness. So, it is likely that these values are well above their true limit. With calendar time portfolios, we get higher abnormal returns using equal weighting than value weighting. This was expected since in our specification tests we saw that equal weighted portfolios gave consistently high and skewed t-statistics compared to value weighted portfolios.

Two factors that strongly influence a firm's decision to participate in a repurchase program are its size and ROA. We saw in Figure 2.1 that a firm in either the highest size decile or the highest ROA decile has about 13% probability of announcing a repurchase. Table 2.6 shows the results of the preceding table applied to size deciles. We group some of the deciles together to obtain a roughly equal number of participating observations in each group. With the RATS method, we see much higher returns for firms in deciles 1 – 8 than for those in the top two deciles, with about 24% compared to 15.6% over 60-months. The opposite is true when using calendar time portfolios. With value weighting, we get 0.39% per month and highly significant returns for firms in deciles 9 – 10, 0.2% per month and not very significant returns for deciles 6 – 8, and no abnormal returns for firms in deciles 1 – 5. Table 2.7 groups firms into deciles of ROA. Similar to the earlier table, we see that a large contribution to abnormal returns comes from firms in the top two ROA deciles.

2.3.3.2 LBT Control Firm

Table 2.8 shows the estimates of buy-and-hold abnormal return obtained usign LBT's control firm. The abnormal returns estimates for the full sample here are much lower than those obtained using the Fama-French regression. Because of delisting repurchasing and control firms, however, we see a large loss in observations from 15% in 12-months to more than 50% in 60-months. This loss creates a concern regarding the efficiency of the procedure. Upon omitting repurchasing and control stocks with pre-announcement price not greater than \$5,[3] neither the estimate nor the significance of the returns changes much. Looking into the size and ROA decile groups gives us interesting results. Within the size decile groups, both the low and mid size groups contribute significantly to 12-month returns. For 36- and 60-month returns, however,

[3] Omitting only event firm stocks creates a negative bias.

while the estimate of the abnormal returns is high, the t-statistic is very low, hinting at the possible inefficiency of using only one control firm. We get a clearer picture from the ROA decile groups. Here, it is clear that most of the positive abnormal returns of repurchasing firms over long horizons can be attributed to the low-ROA firms.

2.3.3.3 Propensity Score Matching

In our analysis thus far, we have utilized propensity score matching in three related ways. We initially set up a caliper of radius 0.01 s.d.[4] and selected all firms within this radius as our initial sample. In case we find no match, we expand the radius to 0.02 s.d. and select the closest match in terms of propensity score. The *All in Caliper* method uses all the firms in this set as controls. From this set, we then chose one or more neighbors closest to the event firm in terms of either the propensity score distance or the Mahalanobis metric. We apply these methods to calculate abnormal returns for firms in our repurchase sample. Since *Size*, *Book-Market*, and *ROA* play a much more significant role in determining participation (as seen in the logistic regression in Table 2.3), we only include these variables in our calculation of the Mahalanobis distance. We show the results of using up to 3 nearest neighbors using either metric in Table 2.9. The sample includes event and control firm stocks whose price at the month-end prior to announcement is at least $5. We see here that there is a substantial difference between the two metrics. In both cases, increasing the number of neighbors leads to an increase in the t-statistic for 36- and 60-months, indicating

[4] Standard deviations of the predicted probabilities from the logistic regression.

an improvement in efficiency. With three neighbors, the propensity score metric shows 3.3%, 6.6%, and 17.7% abnormal returns over 12-, 36-, and 60-months, all significant at the 0.1% level. When all controls within the caliper are included, the 60-month return becomes 10.8%. This decrease in abnormal returns with higher number of neighbors was also seen and discussed in the specification tests in Section 1.5.[5] Compared to the propensity score metric, the abnormal returns obtained using the Mahalanobis metric are lower. Using three nearest neighbors, we get 2.2%, 4.8%, and 10.1% for 12-. 36-, and 60-months respectively, significant at about the 5% level. In this case, the effect of the negative skewness of abnormal returns with increasing neighbors seems to be absent. To explore this issue further, we tabulate the values for abnormal returns using the Mahalanobis metric with one to ten neighbors in Table 2.10. For 60-month returns, we see that even as an increasing number of neighbors are added to the control portfolio, the value of the abnormal returns doesn't change beyond the 9–10% interval. This behavior is in stark contrast to what we evidenced earlier, and also very different from the results we get using the propensity score metric. Since the neighbors in both cases are picked from the same set of controls within the caliper, the difference in results must caused by the order in which neighbors are added. The Mahalanobis metric, therefore, appears to be more successful in identifying good controls.

We now compare the quality of these matches amongst themselves and against LBT's control firm match. To do this, we calculate the difference in the six firm

[5]In unreported results, we found that the abnormal returns over 60-months decreases steadily from 17.7% with 3 neighbors to 12.3% with 10 neighbors, leading more credibility to our earlier claims about potential problems associated with using large number of neighbors.

characteristics used to predict propensity scores between each repurchasing firm and its corresponding control firm(s) using each method. We use three nearest neighbors where applicable. We then map the distribution of these differences and plot the values at decile breakpoints. These plots are in Fig 2.2–Fig 2.7. For instance, in Figure 2.2 we see that at the 1st percentile, the difference in size between a repurchasing and its control firm using LBT's method is -0.5 standard deviations of size.

Figure 2.2 shows the difference in size. As expected, LBT's method controls explicitly for size and is thus very well matched. The other methods do quite well and the only disparities we see are at the 1% level. In Figure 2.3, LBT's method again does very well since it incorporates book-market explicitly as well. The Mahalanobis metric also controls well for book-market, with the difference being only about 0.15 s.d. at the 10th and 90th percentiles. In Figure 2.4, the Mahalanobis metric matches leverage better than the other three. Mahalanobis metric also controls very well for ROA which is one of the factors used to calculate the Mahalanobis distance. This can be seen in Figure 2.5. We also see that LBT's method has issues with matching for ROA at the extreme percentiles. Figure 2.6 shows there's not much difference in matching for cash amongst the four methods. Figure 2.7 shows another scenario where the Mahalanobis metric does much better than the other three methods.

Therefore, the initial caliper extracts all the benefit of using propensity scores. In almost every instance in the figures discussed above, there is no material difference between the set of matches picked out by the caliper and the three nearest neighbors chosen in terms of propensity scores. On the other hand, if the three nearest neighbors are chosen based on the Mahalanobis distance, we see a tangible improvement in the quality of matches. For further analysis, therefore, we strongly recommend using the Mahalanobis metric to select three nearest neighbors and calculate abnormal returns.

Next, we break the repurchase sample into groups of deciles of size and ROA and calculate abnormal returns within each group. The results are in Table 2.11. We see that the abnormal returns observed on the full sample are caused almost exclusively by firms in the lower size and ROA decile groups. The ROA decile groups provide the clearest picture, with zero abnormal returns over 36- and 60-months for firms in the top two deciles. This is also evidence of the matching capability of the Mahalanobis metric.

2.4 Conclusions

We saw definitive evidence showing that the method used to calculate long-run abnormal returns plays an overwhelming role in the estimates we obtain. Using Fama-French regressions, we saw hgh abnormal returns over every time horizon. Ibbotson's RATS method gave us an estimate of 21% over 60-months, while with equal weighted calendar time portfolios we get about 0.5% per month over each time horizon. The RATS estimate is lower than that reported by Peyer & Vermaelen (2009), but the calendar time estimate is almost exactly the same. Some difference is expected since their sample is from 1991–2001[6] and they also confirm each announcement in their sample using *LexisNexis*. We showed in our specification tests that both RATS and equal weighted calendar time portfolios result in estimates with high positive skewness. When we change the weighting scheme to value weighting, we observe lower but

[6]Upon restricting our sample to the same time period, the RATS estimate for 3- and 5-year abnormal returns are about 20% and 27%, which is much closer to their result.

still highly statistically significant returns. The difference is about 10 b.p. per month over 5 years. As we'd seen in Table 2.4, even value weighting is positively skewed when it comes to samples clustered on high size and high ROA, a typical characteristic of firms that announce repurchase programs. The value weighting scheme is also likely to be positively biased for the repurchase sample. We confirm this suspicion in Tables 2.6 and 2.7 where we see that the value weighting method reports high abnormal returns for the top decile groups—with 0.7% per month over the first year for firms in the top two ROA deciles. We then used the control firm method proposed by LBT which is less likely to be skewed on the repurchase sample compared to the regression methods. While we do obtain lower estimates of abnormal returns, the single control firm causes a loss in efficiency. The estimate for 36-months isn't significant at the 5% level, and for the 60-month estimate we end up losing more than half of our sample. This inefficiency adds uncertainty to the estimates provided by this procedure.

We then estimated abnormal returns using propensity matching. In our specification tests, we didn't observe any material difference in the results provided by a propensity score metric and the Mahalanobis metric. However, the estimates or long-run abnormal returns obtained for the repurchase sample are very different. With three neighbors and the propensity score metric, we get 17.7% abnormal return over 60-months compared to about 10% using the Mahalanobis metric. Upon including all the controls in our initial caliper for each event firm, we observe a 10% return as well and also notice that as we increase the number of neighbors, the propensity score metric estimate decreases. This, of course, is symptomatic of methods using multiple control firms, as was discusssed earlier. The Mahalanobis metric, while exhibiting the same symptoms in our specification tests, is a lot more consistent on the repurchase sample. Even changing the number of neighbors from 1 to 10 causes only minor fluctuations in the 5-year abnormal return estimate in the range of 9–10%. Further investigations reveal that the quality of matches obtained by matching on the Mahalanobis metric is much better than that obtained using either the propensity score metric or the LBT control firm. Due to this, we achieve higher efficiency without a compromise on bias, and the 60-month estimate with the Mahalanobis metric and 3 control firms is composed of 50% more observations than the single control firm of LBT. We also observe that the abnormal returns can be primarily attributed to firms with lower ROA than the median in a particular year.

We thus conclude that firms announcing open market share repurchase programs do earn a positive abnormal return over 1, 3, and 5 years. The magnitude of this return is a lot lower than what has been previously reported. Firms with high size and high ROA are particularly likely to announce a repurchase program. However, it is the firms in the lower end of the distribution that earn abnormal returns over long horizons. A second note of significance is that we are able to demonstrate the usefulness of the propensity score matching procedure, especially when we employ the right caliper in conjunction with the Mahalanobis metric. We strongly recommend incorporating this procedure in studies concerned with long-run returns.

Table 2.1: Summary Statistics of Repurchasing Firms

This table presents summary statistics for Repurchasing and Non-Repurchasing firms. A firm is classified as "Repurchasing" at December of a particular year if it has at least one repurchase announcement during the next calendar year. Each of the characteristics presented below has been normalized to have zero mean and unit standard deviation. The definitions are: *Size* - Market Value in 1987 Dollars; *Book-Market* - Ratio of book to market values of common stock; *Leverage* Ratio of total liabilities to total assets; *ROA* - Ratio of net income to total assets; *Cash & Short Term Inv.* - Ratio of cash & short term investments to total assets; *Total Asset Turnover* - Ratio of total sales to total assets. The last column shows the t-statistic for the difference in means between Repurchasing and Non-Repurchasing firms. *, ** denote significance levels at 5% and 1% using a two-tailed test.

	Repurchasing Firms				Non-Repurchasing Firms		Mean difference
	Min.	Median	Mean	Max.	Median	Mean	t-statistic
Size	-0.157	-0.12	0.191	31.78	-0.147	-0.016	12.027**
Book-Market	-0.744	-0.235	-0.149	6.102	-0.177	0.013	-26.531**
Leverage	-2.094	0.076	0.084	1.903	0.022	-0.007	7.382**
ROA	-6.137	0.261	0.277	4.763	0.165	-0.023	53.377**
Cash & Short Term Inv.	-0.783	-0.442	-0.078	4.172	-0.432	0.007	-8.003**
Total Asset Turnover	-1.119	-0.133	-0.04	12.15	-0.128	0.003	-3.932**

Table 2.2: Decile Distribution of Repurchasing Firms

This tabble shows the distribution of Repurchasing firms in deciles of the firm characteristics. The decile breakpoints are calculated each year. Numbers shown are in terms of percentage of the total number of repurchasing announcements. A firm is classified as "Repurchasing" at December of a particular year if it has at least one repurchase announcement during the next calendar year. Each of the characteristics presented below has been normalized to have zero mean and unit standard deviation. The definitions are: *Size* - Market Value in 1987 Dollars; *Book-Market* - Ratio of book to market values of common stock; *Leverage* Ratio of total liabilities to total assets; *ROA* - Ratio of net income to total assets; *Cash & Short Term Inv.* - Ratio of cash & short term investments to total assets; *Total Asset Turnover* - Ratio of total sales to total assets; V_{net} = *Size* - *Book-Market* + *ROA*.

	1	2	3	4	5	6	7	8	9	10
Size	2.2	4.7	6.1	8	9.2	9.6	11	12.7	16.1	20.5
Book-Market	7.1	10.9	11.6	11.6	11.7	11.3	11.3	10.6	9.4	4.6
Leverage	8	10.1	10	10.8	9.7	10.6	9	8.1	11.7	12
ROA	1.3	3.9	6	12.6	12.2	10.3	10.8	13.4	14.9	14.6
Cash & Short Term Inv.	7.4	9.8	10.6	11.4	10.9	10.9	10.6	10.2	11.1	7
Total Asset Turnover	11.4	11.3	8.1	8.7	10.5	10.7	10	10.3	9.2	9.8
V_{net}	2	3.7	7.4	8.5	10.4	10.2	11.1	12.6	14.8	19.3

Table 2.3: Logistic Regression predicting Repurchase Announcements

This table presents coefficients and t-statistics for the logistic regression predicting a firm's probability to announce a repurchase program. The dependent variable is defined at the firm-year level and takes on a value of 1 for a repurchasing firm. A firm is classified as "Repurchasing" at December of a particular year if it has at least one repurchase announcement during the next calendar year. Each of the independent variables presented below has been normalized to have zero mean and unit standard deviation. The definitions are: *Size* - Market Value in 1987 Dollars; *Book-Market* - Ratio of book to market values of common stock; *Leverage* Ratio of total liabilities to total assets; *ROA* - Ratio of net income to total assets; *Cash & Short Term Inv.* - Ratio of cash & short term investments to total assets; *Total Asset Turnover* - Ratio of total sales to total assets. *, ** , *** denote significance levels at 5%, 1%, and 0.1% using a two-tailed test.

	Linear Model		Non-Linear Model	
	Estimate	t-statistic	Estimate	t-statistic
$Size$	0.063	8.1***	0.275	13.1***
$Size^2$			-0.012	-8.1***
$Book\_Market$	-0.43	-15.9***	-0.313	-8.8***
$Book\_Market^2$			-0.076	-2.9**
$Leverage$	0.102	6.8***	0.095	6.2***
$Leverage^2$			0.052	3.8***
ROA	0.783	28.9***	0.987	26***
ROA^2			-0.217	-8.6***
$Cash$	-0.04	-2.5*	0.077	3**
$Cash^2$			-0.058	-4.7***
TAT	-0.164	-10.8***	-0.149	-9***
TAT^2			0.002	2.8***

Table 2.4: Specification Tests - V_{net}-Clustered Samples

We select 200 firm-year observations from the highest V_{net} decile and assign each of these an event. For each event, we assign an event month of January. This produces non-random V_{net}-clustered samples. We then calculate long-run abnormal returns using the methods tabulated below. We repeat this exercise 2,000 times and compute the percentage of t-statistics from each method that falls in the 5%, 2.5%, or 0.5% tails on the left and right sides of the corresponding t-distribution. For propensity score matching, we use a caliper of 0.01 standard deviations of the predicted values from the corresponding logistic regression to select our matches from. If this doesn't produce any matches, we use an additional caliper of 0.02 standard deviations from which one match is selected. The variables used to calculate estimated propensity scores are Size, Book-Market, Leverage, Return on Assets, Cash & Cash Equivalents, and Total Asset Turnover and their squares.

		Theoretical CDF (%)					
		0.5	99.5	2.5	97.5	5	95
	12-Month Returns						
		Propensity Score Matching (5 Variables)					
Propensity Score Distance	1 Nearest Neighbor in Caliper	0.30	0.35	2.15	2.75	5.35	4.90
	2 Nearest Neighbors in Caliper	0.75	0.35	3.20	2.10	5.70	4.20
	3 Nearest Neighbors in Caliper	0.70	0.30	2.80	1.90	6.45	4.20
	4 Nearest Neighbors in Caliper	0.60	0.40	3.15	2.35	6.40	4.05
	5 Nearest Neighbors in Caliper	0.75	0.40	3.05	2.25	6.25	4.05
	All in Caliper	0.55	0.50	3.20	2.10	6.20	4.60
Mahalanobis Distance	1 Nearest Neighbor in Caliper	0.35	0.60	2.20	3.05	4.25	6.60
	2 Nearest Neighbors in Caliper	0.40	0.60	2.00	3.15	4.65	6.10

Continued on Next Page...

Table 2.4 – Continued

		Theoretical CDF (%)					
		0.5	99.5	2.5	97.5	5	95
	3 Nearest Neighbors in Caliper	0.45	0.60	2.35	2.80	4.80	6.05
	4 Nearest Neighbors in Caliper	0.35	0.50	2.40	2.75	5.30	5.55
	5 Nearest Neighbors in Caliper	0.40	0.60	2.75	2.65	5.30	5.65
		Conventional Methods					
Control Firm	LBT Control Firm	0.15	1.35	0.95	5.20	2.55	8.90
	5 Random Control Firms	0.15	1.00	1.05	5.15	1.80	10.25
	Ibbotson's RATS	0.05	1.85	0.15	7.10	0.45	13.90
Calendar Time Portfolios	Equal Weghted w/ FF 4-factor	0.05	1.90	0.40	9.25	0.80	16.50
	Value Weighted w/ FF 4-factor	0.05	0.80	0.40	4.35	0.95	9.40
36-Month Returns							
		Propensity Score Matching (5 Variables)					
Propensity Score Distance	1 Nearest Neighbor in Caliper	0.20	0.75	2.35	3.00	4.75	5.95
	2 Nearest Neighbors in Caliper	0.40	0.50	2.95	2.45	6.00	4.70
	3 Nearest Neighbors in Caliper	0.60	0.25	3.40	2.05	6.30	4.25

Continued on Next Page...

Table 2.4 – Continued

		Theoretical CDF (%)					
		0.5	99.5	2.5	97.5	5	95
	4 Nearest Neighbors in Caliper	0.75	0.20	3.65	1.95	6.40	4.15
	5 Nearest Neighbors in Caliper	0.70	0.20	3.85	1.90	6.50	4.10
	All in Caliper	1.05	0.05	4.50	1.45	7.55	4.00
Mahalanobis Distance	1 Nearest Neighbor in Caliper	0.20	0.50	1.95	2.85	4.15	6.55
	2 Nearest Neighbors in Caliper	0.50	0.30	2.10	2.90	4.65	6.75
	3 Nearest Neighbors in Caliper	0.45	0.20	2.35	2.55	4.80	5.80
	4 Nearest Neighbors in Caliper	0.45	0.30	2.70	2.45	4.80	5.65
	5 Nearest Neighbors in Caliper	0.35	0.30	2.45	2.45	4.95	5.75
		Conventional Methods					
Control Firm	LBT Control Firm	0.25	0.95	0.90	3.60	2.15	8.20
	5 Random Control Firms	0.20	0.25	2.05	2.35	4.20	4.90
	Ibbotson's RATS	0.00	2.90	0.05	12.60	0.10	22.75
Calendar Time Portfolios	Equal Weghted w/ FF 4-factor	0.00	3.85	0.05	14.45	0.10	25.10
	Value Weighted w/ FF 4-factor	0.05	2.00	0.30	8.75	0.55	15.10

Continued on Next Page...

Table 2.4 – Continued

		Theoretical CDF (%)					
		0.5	99.5	2.5	97.5	5	95
60-Month Returns							
Propensity Score Matching (5 Variables)							
Propensity Score Distance	1 Nearest Neighbor in Caliper	0.45	0.40	2.30	2.50	4.80	5.15
	2 Nearest Neighbors in Caliper	0.50	0.60	2.40	1.75	5.30	4.50
	3 Nearest Neighbors in Caliper	0.65	0.20	3.05	1.70	5.95	3.65
	4 Nearest Neighbors in Caliper	0.70	0.25	2.85	1.50	5.95	3.20
	5 Nearest Neighbors in Caliper	0.75	0.15	2.95	1.35	6.00	3.20
	All in Caliper	0.95	0.10	3.90	0.85	7.35	2.05
Mahalanobis Distance	1 Nearest Neighbor in Caliper	0.15	0.75	1.65	3.25	3.80	5.85
	2 Nearest Neighbors in Caliper	0.45	0.45	1.60	2.65	3.60	5.20
	3 Nearest Neighbors in Caliper	0.35	0.40	2.10	2.30	4.15	4.80
	4 Nearest Neighbors in Caliper	0.45	0.35	2.10	2.10	4.65	4.80
	5 Nearest Neighbors in Caliper	0.50	0.35	2.40	1.80	5.05	4.60
Conventional Methods							

Continued on Next Page...

Table 2.4 – Continued

		Theoretical CDF (%)					
		0.5	99.5	2.5	97.5	5	95
Control Firm	LBT Control Firm	0.10	0.75	1.05	2.90	2.25	6.35
	5 Random Control Firms	0.60	0.25	2.40	1.60	5.30	4.00
	Ibbotson's RATS	0.00	9.70	0.00	26.90	0.10	39.90
Calendar Time Portfolios	Equal Weghted w/ FF 4-factor	0.00	19.60	0.05	41.25	0.10	54.60
	Value Weighted w/ FF 4-factor	0.40	4.80	0.90	15.20	1.95	23.75

Table 2.5: Repurchase Returns - Fama French Regressions

This table presents estimates of long-run abnormal returns following repurchase announcements using the RATS method and calendar time (C-Time) equal/value weighted portfolios. Abnormal returns are calculated starting from the month following the announcement. Panel A includes the full sample while Panel B considers a subset of the entire sample comprising stocks whose price was higher than $5 at the end of the month prior to announcement. All returns estimates are in percentages. The calendar time values are monthly returns. *, ** , *** denote significance levels at 5%, 1%, and 0.1% using a two-tailed test.

	12-Months		36-Months		60-Months	
	Estimate	t-statistic	Estimate	t-statistic	Estimate	t-statistic
			Panel A: Full Sample (7,377 Obs.)			
RATS	6.9	11.7***	17.5	16.25***	23.6	16.6***
C-Time (EW)	0.63	5.72***	0.57	5.93***	0.51	5.75***
C-Time (VW)	0.45	4.1***	0.37	4.3***	0.37	4.4***
			Panel B: Prior Price > $5 (6,663 Obs.)			
RATS	5.8	10.8***	15.3	15.1***	21	15.5***
C-Time (EW)	0.51	4.6***	0.49	5.1***	0.46	5.2***
C-Time (VW)	0.45	4.1***	0.36	4.3***	0.36	4.4***

Table 2.6: Repurchase Returns in Size Deciles - Fama French Regressions

This table presents estimates of long-run abnormal returns following repurchase announcements using the RATS method and calendar time (C-Time) equal/value weighted portfolios within size deciles. Panels A, B, and C include firms in size deciles 9–10, 6–8, and 1–5, respectively. Abnormal returns are calculated starting from the month following the announcement. The sample comprises stocks whose price was higher than $5 at the end of the month prior to announcement. All returns estimates are in percentages. The calendar time values are monthly returns. *, ** , *** denote significance levels at 5%, 1%, and 0.1% using a two-tailed test.

	12-Months		36-Months		60-Months	
	Estimate	t-statistic	Estimate	t-statistic	Estimate	t-statistic
			Panel A: Size Deciles 9 – 10 (2,690 Obs.)			
RATS	5.8	7.7***	12.2	9***	15.6	8.9***
C-Time (EW)	0.55	4.7***	0.47	4.7***	0.47	5***
C-Time (VW)	0.45	4***	0.37	4.3***	0.39	4.8***
			Panel B: Size Deciles 6 – 8 (2,359 Obs.)			
RATS	6.8	6.8***	17.7	9.5***	24.1	9.6***
C-Time (EW)	0.34	2.1*	0.37	2.7**	0.36	2.7**
C-Time (VW)	0.31	1.98*	0.3	2.3**	0.2	1.6
			Panel C: Size Deciles 1 – 5 (1,614 Obs.)			
RATS	4.8	4.3***	16.3	7.1***	24.4	7.6***
C-Time (EW)	0.23	1.2	0.32	2.1**	0.26	1.9*
C-Time (VW)	0.16	0.8	0.18	1.1	0	0.2

Table 2.7: Repurchase Returns in ROA Deciles - Fama French Regressions

This table presents estimates of long-run abnormal returns following repurchase announcements using the RATS method and calendar time (C-Time) equal/value weighted portfolios within roa deciles. Panels A, B, and C include firms in size deciles 9–10, 6–8, and 1–5, respectively. Abnormal returns are calculated starting from the month following the announcement. The sample comprises stocks whose price was higher than $5 at the end of the month prior to announcement. All returns estimates are in percentages. The calendar time values are monthly returns. *, ** , *** denote significance levels at 5%, 1%, and 0.1% using a two-tailed test.

	12-Months		36-Months		60-Months	
	Estimate	t-statistic	Estimate	t-statistic	Estimate	t-statistic
Panel A: ROA Deciles 9 – 10 (2,041 Obs.)						
RATS	8	7.57***	19.4	10.05***	24.3	9.6***
C-Time (EW)	0.61	3.96***	0.55	4.33***	0.52	4.39***
C-Time (VW)	0.7	3.55***	0.48	3.41***	0.5	3.95***
Panel B: ROA Deciles 6 – 8 (2,343 Obs.)						
RATS	3.9	4.34***	10	5.98***	16.3	7.29***
C-Time (EW)	0.35	2.76**	0.31	3.11**	0.33	3.58***
C-Time (VW)	0.28	1.86	0.24	2.39**	0.17	1.79
Panel C: ROA Deciles 1 – 5 (2,279 Obs.)						
RATS	5.9	6.8***	17.1	10.15***	23	10.1***
C-Time (EW)	0.4	2.85**	0.48	4.06***	0.44	3.78***
C-Time (VW)	0.01	0.51	0.35	2.85**	0.13	2.66**

Table 2.8: Repurchase Returns - LBT Control Firm

This table presents estimates of long-run abnormal returns following repurchase announcements using the LBT control firm. Results are shown for the full sample and the subsample comprising stocks (both repurchasing and control firms) whose price was higher than $5 at the end of the month prior to announcement. Within this subsample we also calculate abnormal returns within groups of size and ROA deciles. The deciles are grouped in 9–10, 6–8, and 1–5. Abnormal returns are calculated starting from the month following the announcement. All returns estimates are in percentages. *, ** , *** denote significance levels at 5%, 1%, and 0.1% using a two-tailed test.

	12-Months			36-Months			60-Months		
	Est.	t-stat	Obs.	Est.	t-stat	Obs.	Est.	t-stat	Obs.
Full Sample	3.2	2.82**	6,339	6.1	1.94	4,690	10.8	2.11*	3,598
Price > $5	3.9	3.81***	5,332	5.2	1.88	4,068	10.3	2.17*	3,196
Size: 9 – 10	1.8	1.45	2,453	3.5	1.02	2,075	4.7	0.8	1,752
Size: 6 – 8	5.4	2.61**	1,931	7.3	1.30	1,380	19.9	2.33*	1,057
Size: 1 – 5	6.4	2.79**	946	6.5	0.93	611	9.4	0.55	385
ROA: 9 – 10	6.2	2.97**	1,726	-1.9	-0.30	1,382	6.9	0.82	1,134
ROA: 6 – 8	1.7	1.08	1,897	3.8	0.94	1,446	7.6	0.89	1,167
ROA: 1 – 5	4.1	2.45*	1,707	15	4.11***	1,218	18	2.58**	893

Table 2.9: Repurchase Returns – Propensity Score Matching

This table presents estimates of long-run abnormal returns following repurchase announcements using propensity score matching. We initially start with a caliper of 0.01 standard deviations of the predicted values from the corresponding logistic regression to select our matches from. If this doesn't produce any matches, we use an additional caliper of 0.02 standard deviations from which one match is selected. The variables used to calculate estimated propensity scores are Size, Book-Market, Leverage, Return on Assets, Cash & Cash Equivalents, and Total Asset Turnover and their squares. Abnormal returns are calculated starting from the month following the announcement. The sample comprises stocks (both repurchasing and control) whose price was higher than $5 at the end of the month prior to announcement. All returns estimates are in percentages. The calendar time values are monthly returns. *, ** , *** denote significance levels at 5%, 1%, and 0.1% using a two-tailed test.

	12-Months			36-Months			60-Months		
	Est.	t-stat	Obs.	Est.	t-stat	Obs.	Est.	t-stat	Obs.
Propensity Score: 1NN	2.2	2.01*	5730	7.4	2.97**	4261	13.7	3.4***	3306
Propensity Score: 2NN	3.1	3.41***	6190	7.4	3.70***	5087	17.4	4.86***	4228
Propensity Score: 3NN	3.3	3.94***	6233	6.6	3.33***	5264	17.7	5.33***	4483
All in Caliper	3.2	4.38***	6235	6.3	3.93***	5320	10.8	3.56***	4610
Mahalanobis Metric: 1NN	2.3	2.32*	5761	3.0	1.11	4311	7.9	1.29	3345
Mahalanobis Metric: 2NN	1.7	2.00*	6180	4.2	2.00*	5090	9.1	2.31*	4212
Mahalanobis Metric: 3NN	2.2	2.72**	6231	4.8	2.54*	5260	10.1	2.71**	4459

Table 2.10: Repurchase Returns - Propensity Score Matching w/ Mahalanobis Metric

This table presents estimates of long-run abnormal returns following repurchase announcements using propensity score matching with Mahalanobis metric. We initially start with a caliper of 0.01 standard deviations of the predicted values from the corresponding logistic regression to select our matches from. If this doesn't produce any matches, we use an additional caliper of 0.02 standard deviations from which one match is selected. The variables used to calculate estimated propensity scores are Size, Book-Market, Leverage, Return on Assets, Cash & Cash Equivalents, and Total Asset Turnover and their squares. Abnormal returns are calculated starting from the month following the announcement. The sample comprises stocks (both repurchasing and control) whose price was higher than $5 at the end of the month prior to announcement. All returns estimates are in percentages. The calendar time values are monthly returns. *, ** , *** denote significance levels at 5%, 1%, and 0.1% using a two-tailed test.

	12-Months			36-Months			60-Months		
Neighbors	Est.	t-stat	Obs.	Est.	t-stat	Obs.	Est.	t-stat	Obs.
1	2.3	2.32*	5761	3.0	1.11	4311	7.9	1.29	3345
2	1.7	2.00*	6180	4.2	2.00*	5090	9.1	2.31*	4212
3	2.2	2.72**	6231	4.8	2.54*	5260	10.1	2.71**	4459
4	2.4	3.07**	6235	4.8	2.64**	5302	9.8	2.52*	4538
5	2.3	3.03**	6235	4.5	2.59**	5314	10.0	2.66**	4581
6	2.2	3.01**	6235	4.7	2.76**	5319	9.0	2.50*	4602
7	2.3	3.15**	6235	5.0	2.98**	5320	9.6	2.83**	4606
8	2.3	3.11**	6235	5.0	3.02**	5320	10.1	3.05**	4609

Continued on Next Page. . .

Table 2.10 – Continued

	12-Months			36-Months			60-Months		
	Est.	t-stat	Obs.	Est.	t-stat	Obs.	Est.	t-stat	Obs.
9	2.5	3.36^{***}	6235	5.2	3.14^{**}	5320	8.9	2.58^{**}	4610
10	2.6	3.51^{***}	6235	5.3	3.24^{**}	5320	9.0	2.67^{**}	4610

Table 2.11: Mahalanobis Metric Returns - Size & ROA Deciles

This table presents estimates of long-run abnormal returns following repurchase announcements using propensity score matching with Mahalanobis metric within groups of size and ROA deciles. We initially start with a caliper of 0.01 standard deviations of the predicted values from the corresponding logistic regression to select our matches from. If this doesn't produce any matches, we use an additional caliper of 0.02 standard deviations from which one match is selected. The variables used to calculate estimated propensity scores are Size, Book-Market, Leverage, Return on Assets, Cash & Cash Equivalents, and Total Asset Turnover and their squares. Abnormal returns are calculated starting from the month following the announcement. The sample comprises stocks (both repurchasing and control) whose price was higher than $5 at the end of the month prior to announcement. All returns estimates are in percentages. The calendar time values are monthly returns. *, ** , *** denote significance levels at 5%, 1%, and 0.1% using a two-tailed test.

	12-Months			36-Months			60-Months		
	Est.	t-stat	Obs.	Est.	t-stat	Obs.	Est.	t-stat	Obs.
Size: 9 – 10	1.2	1.11	2461	3.7	1.55	2221	3.2	0.74	1981
Size: 6 – 8	3.1	2.01*	2240	2.9	0.87	1865	13.8	1.87	1597
Size: 1 – 5	2.4	1.53	1528	10.0	2.03*	1172	18.8	2.09*	879
ROA: 9 – 10	2.1	1.16	1896	-0.1	-0.02	1679	0.0	0.00	1483
ROA: 6 – 8	1.4	1.13	2197	4.5	1.36	1884	13.6	2.35*	1639
ROA: 1 – 5	3.0	2.76**	2136	10.1	3.61***	1695	17.0	3.39***	1335

Figure 2.1: Predicted Probability by Decile

This figure shows the predicted probabilities obtained from the logistic regression for repurchase decisions plotted by deciles of the independent variables shown. The horizontal axis shows the decile values. The vertical axis shows the predicted probabilities. For each variable-decile value, we plot the predicted probabilities based on the mean values of all the independent variables in that variable-decile. The horizontal line in the middle shows the baseline probability, i.e. the predicted probability for a firm with all its variables at 0. The bold line shows the predicted probabilities according to decile breakpoints of V_{net}.

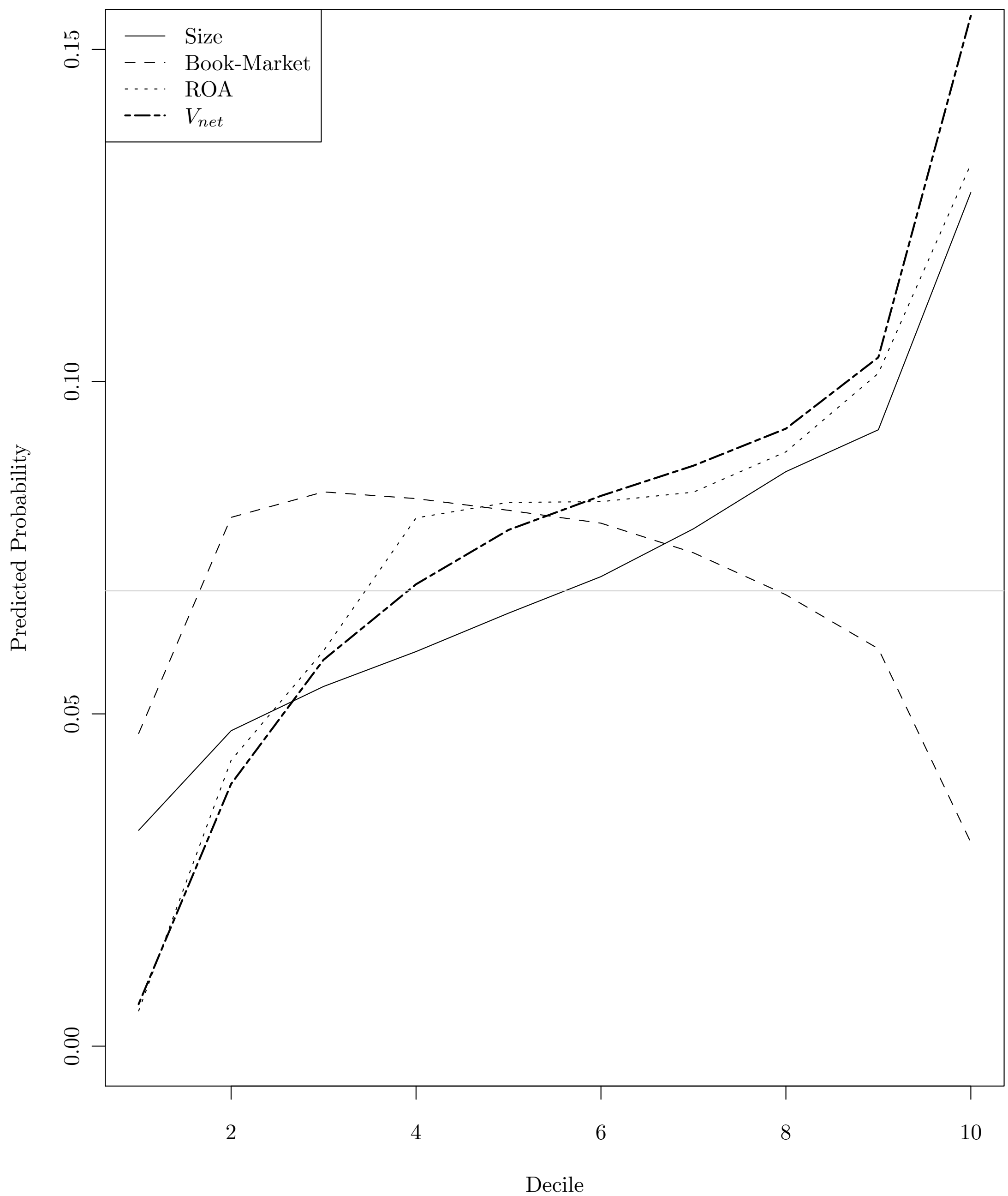


Figure 2.2: Repurchasing & Control Firm(s): Size Difference

We calulate the difference between the size of repurchasing firms and their matched control firm(s) for each firm announcing a repurchase. This figure shows the distribution of the difference using various methods. The horizontal axis shows the percentile while the vertical axis shows the difference in units of standard deviations of the underlying variable.

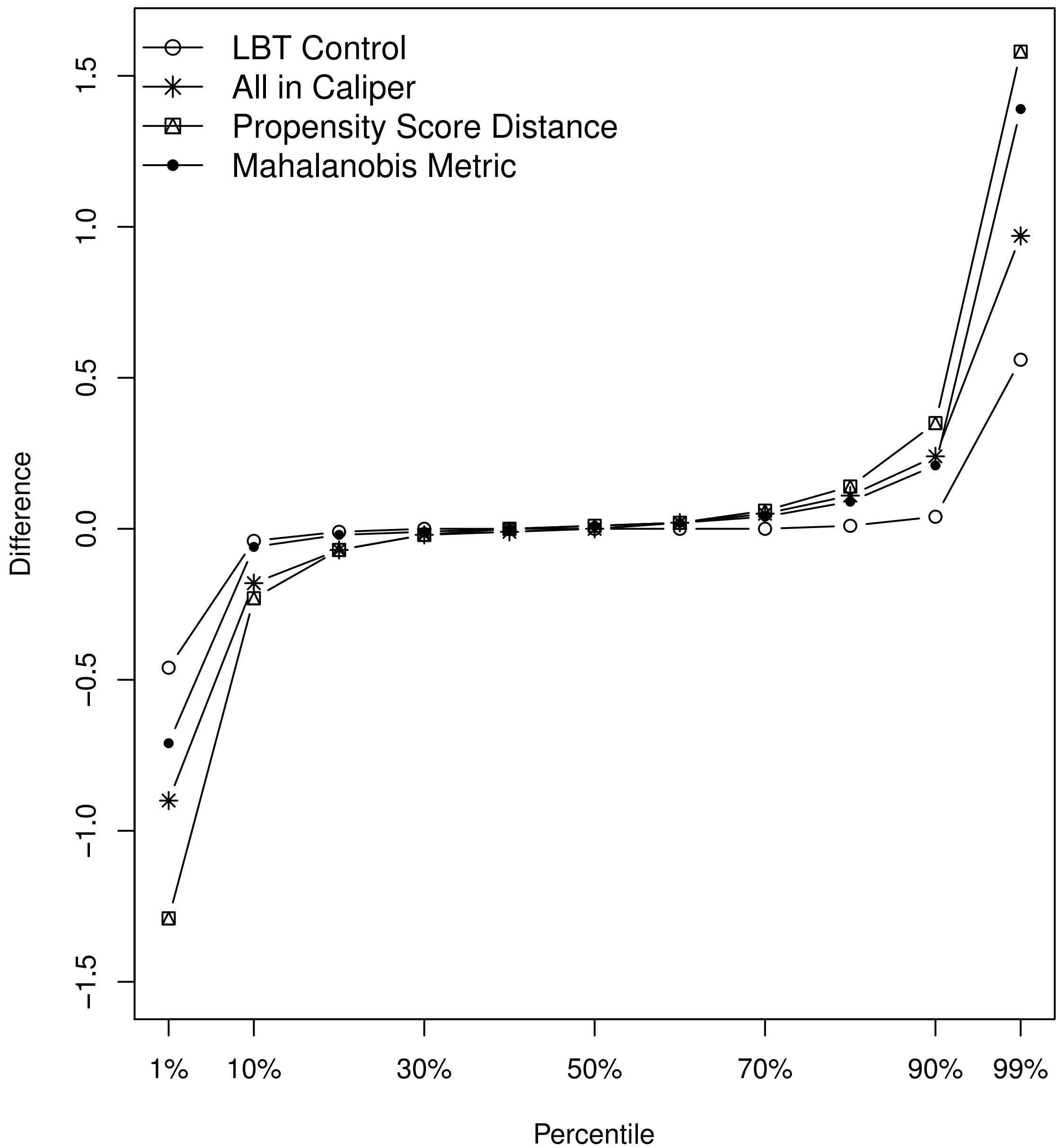


Figure 2.3: Repurchasing & Control Firm(s): Book_Market Difference

We calulate the difference between the book_market of repurchasing firms and their matched control firm(s) for each firm announcing a repurchase. This figure shows the distribution of the difference using various methods. The horizontal axis shows the percentile while the vertical axis shows the difference in units of standard deviations of the underlying variable.

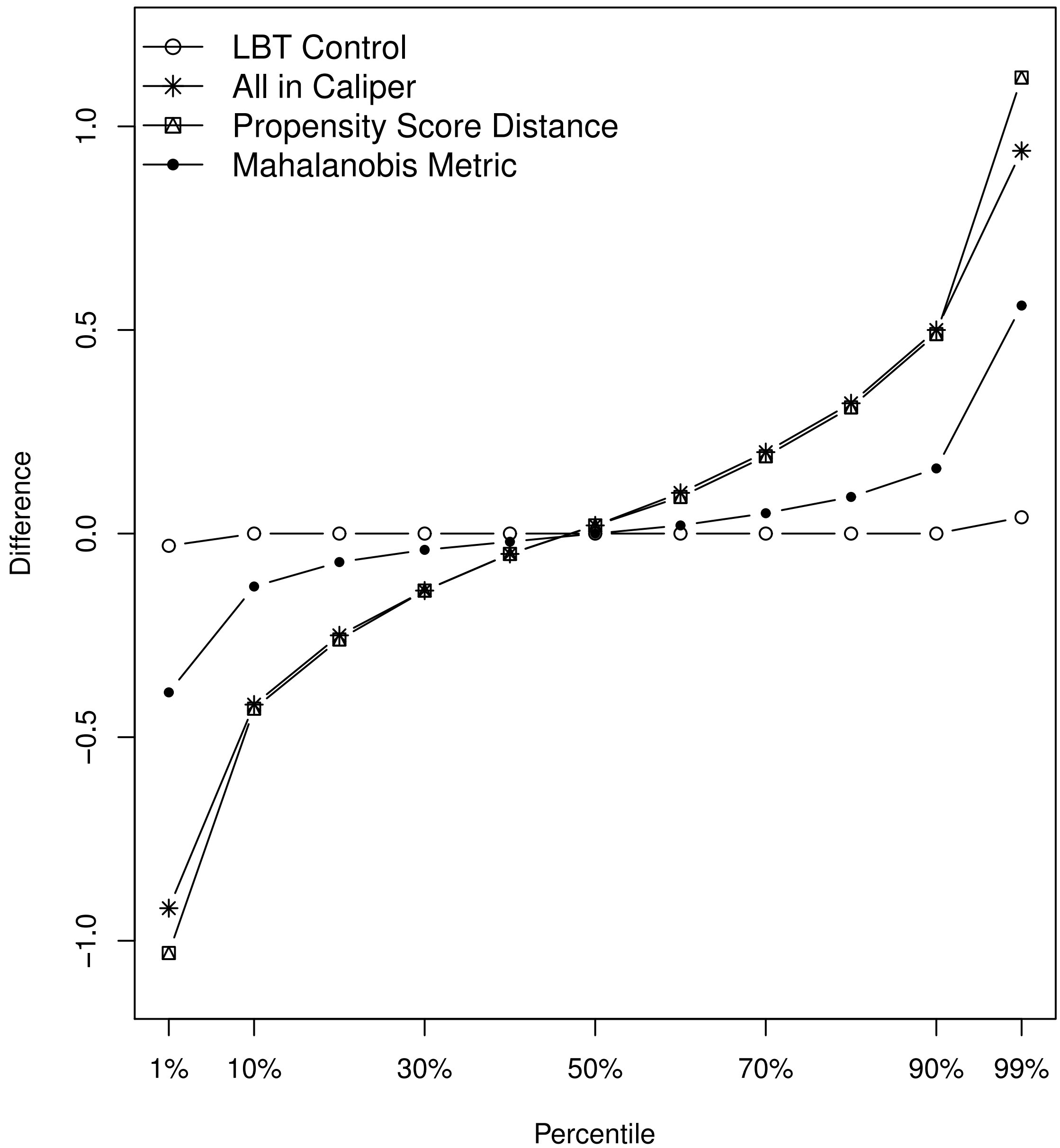


Figure 2.4: Repurchasing & Control Firm(s): Leverage Difference

We calulate the difference between the leverage of repurchasing firms and their matched control firm(s) for each firm announcing a repurchase. This figure shows the distribution of the difference using various methods. The horizontal axis shows the percentile while the vertical axis shows the difference in units of standard deviations of the underlying variable.

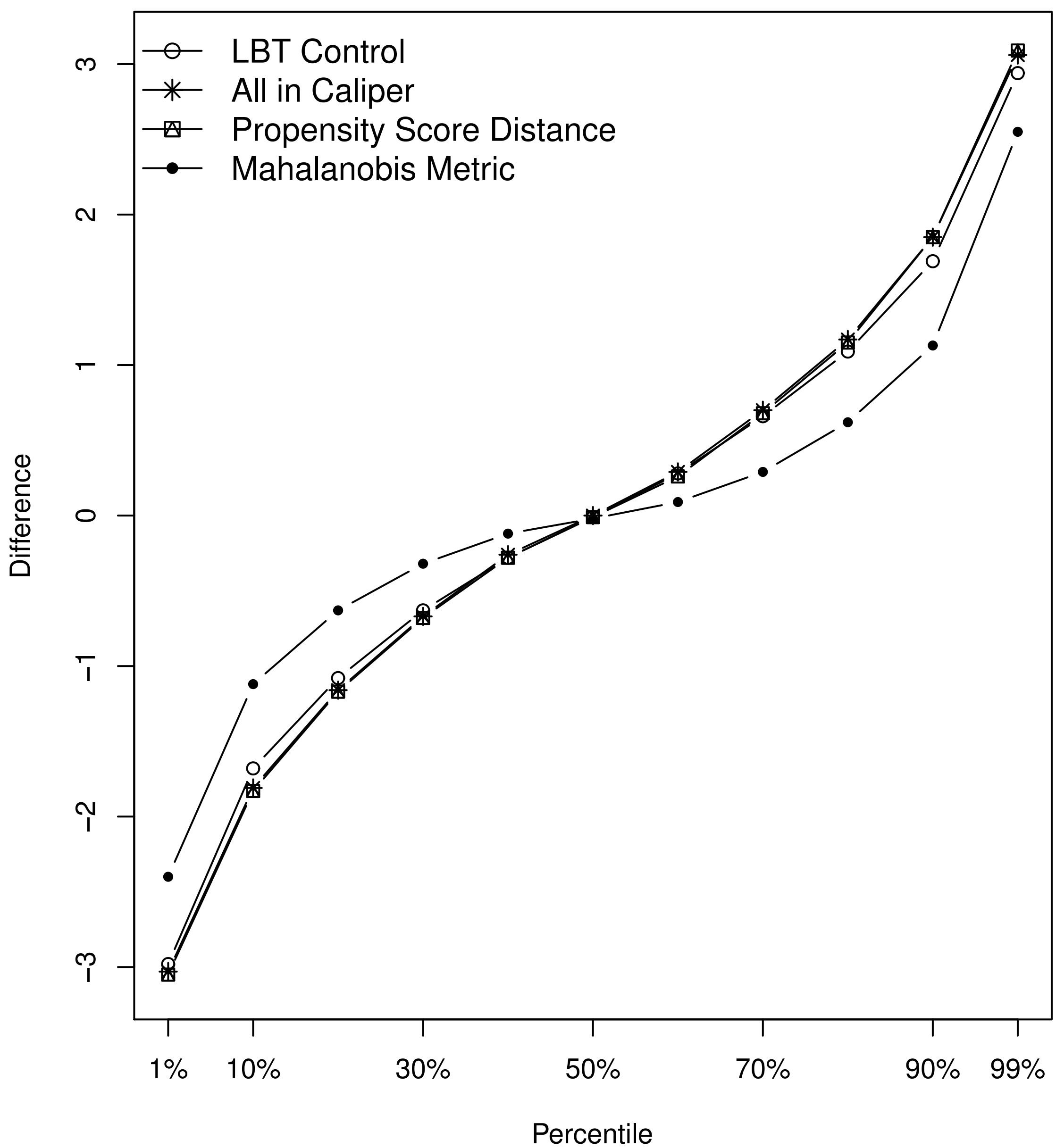


Figure 2.5: Repurchasing & Control Firm(s): ROA Difference

We calulate the difference between the ROA of repurchasing firms and their matched control firm(s) for each firm announcing a repurchase. This figure shows the distribution of the difference using various methods. The horizontal axis shows the percentile while the vertical axis shows the difference in units of standard deviations of the underlying variable.

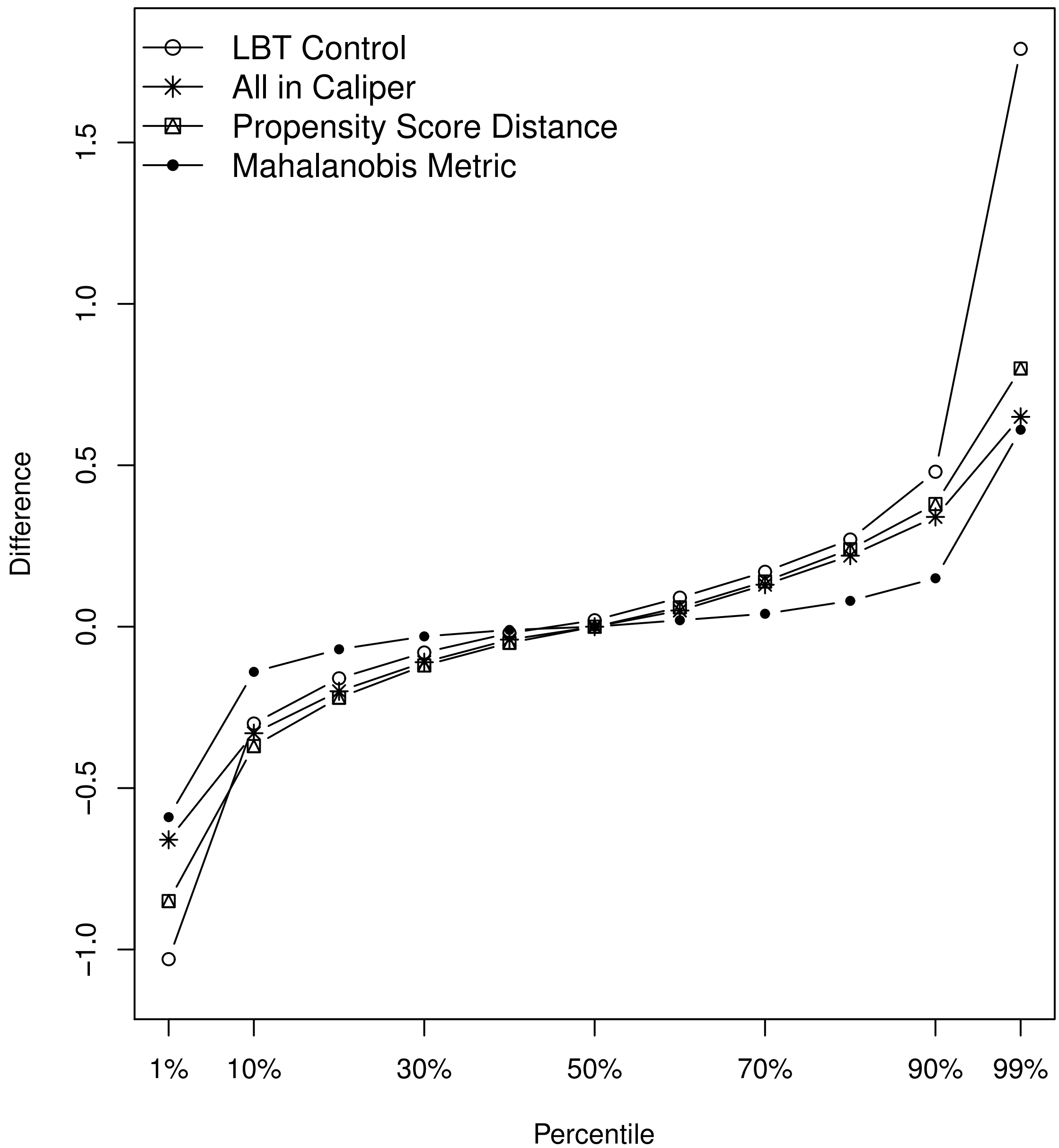


Figure 2.6: Repurchasing & Control Firm(s): Cash Difference

We calulate the difference between the cash & cash equivalents of repurchasing firms and their matched control firm(s) for each firm announcing a repurchase. This figure shows the distribution of the difference using various methods. The horizontal axis shows the percentile while the vertical axis shows the difference in units of standard deviations of the underlying variable.

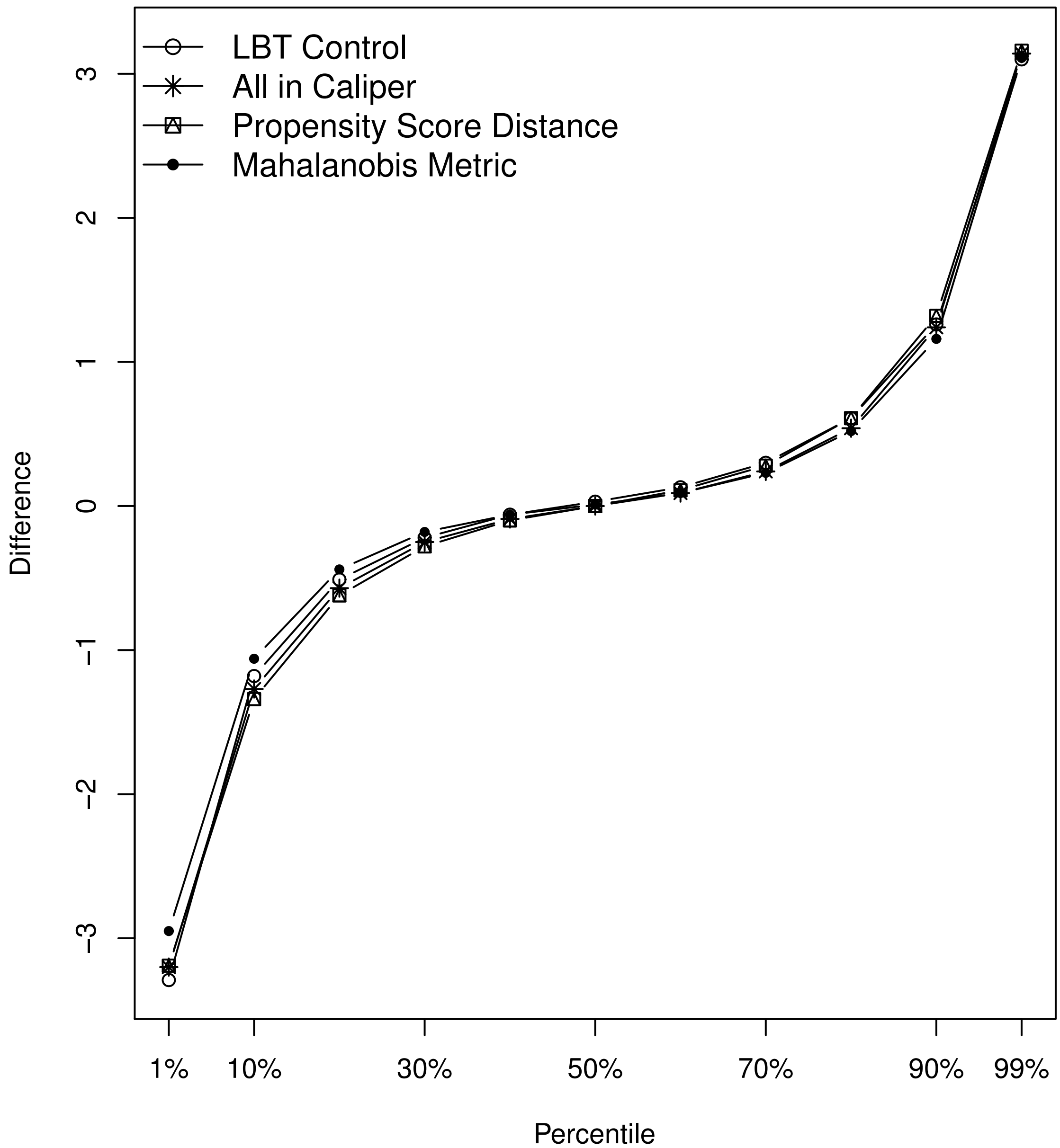
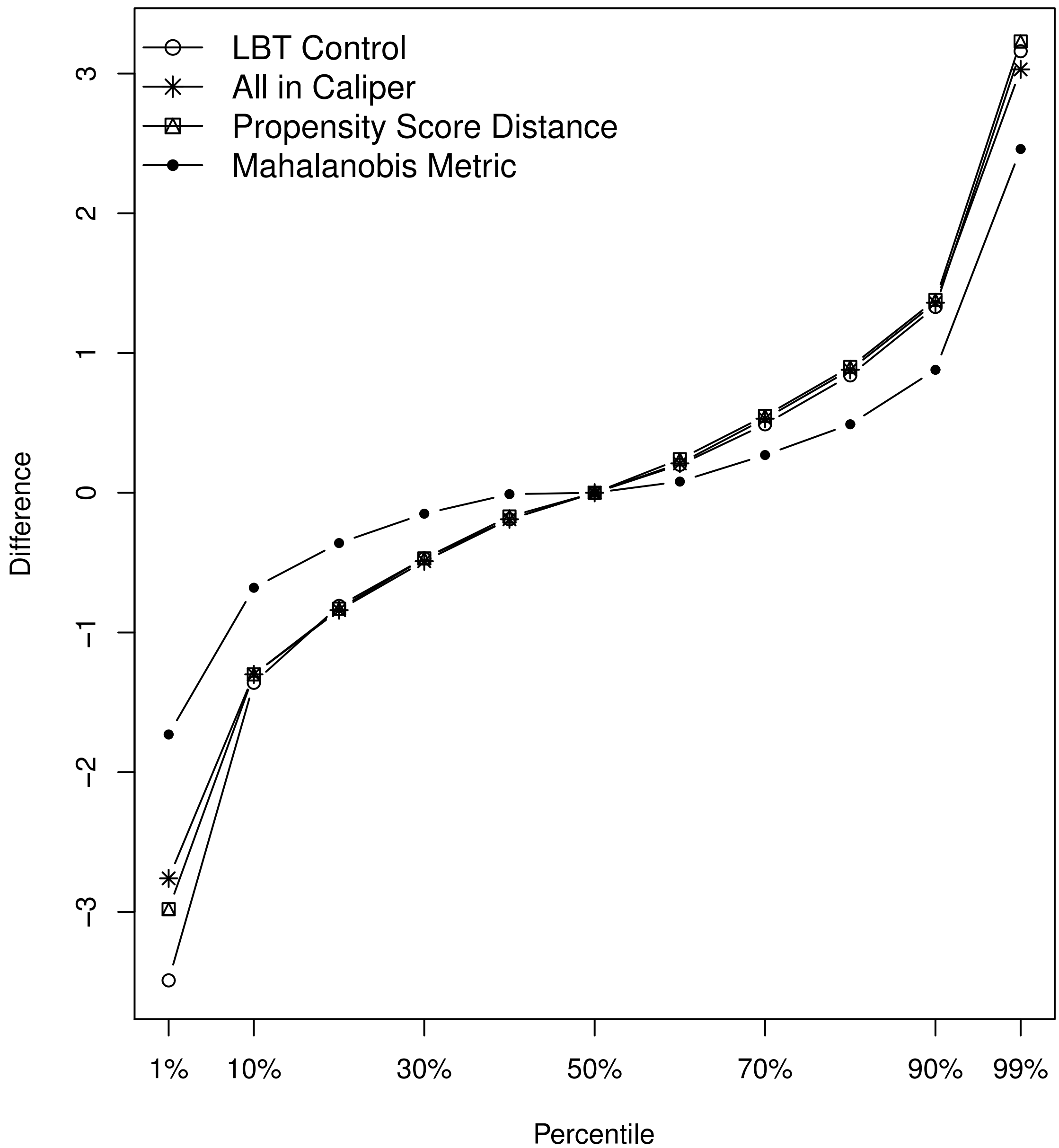


Figure 2.7: Repurchasing & Control Firm(s): TAT Difference

We calulate the difference between the tat of repurchasing firms and their matched control firm(s) for each firm announcing a repurchase. This figure shows the distribution of the difference using various methods. The horizontal axis shows the percentile while the vertical axis shows the difference in units of standard deviations of the underlying variable.



Bibliography

R. Aggarwal, I. Erel, R. M. Stulz, and R. Williamson. Do U.S. Firms Have the Best Corporate Governance? A Cross-Country Examination of the Relation Between Corporate Governance and Shareholder Wealth. *ECGI - Finance Working Paper No. 145/2007*, Jan 2007. http://ssrn.com/abstract=954169.

R. Aggarwal, I. Erel, R. Stulz, and R. Williamson. Differences in Governance Practices between U. S. and Foreign Firms: Measurement, Causes, and Consequences. *The Review of Financial Studies*, 22(8):pp. 3131–3169, Aug 2009.

S. Ahn and M. D. Walker. Corporate Governance and the Spinoff Decision. *Journal of Corporate Finance*, 13(1), Mar 2007.

W. Z. Billewicz. The Efficiency of Matched Samples: An Empirical Investigation. *Biometrica*, 21, Sep 1965.

A. Brav, J. R. Graham, C. R. Harvey, and R. Michaely. Payout Policy in the 21st Century. *Journal of Financial Economics*, 77(3):483 – 527, 2005.

K. Chan, D. L. Ikenberry, and I. Lee. Do managers time the market? Evidence from Open-Market Share Repurchases. *Journal of Banking & Finance*, 31(9):2673 – 2694, 2007.

F. S. Chapin. Design for Social Experiments. *American Sociological Review*, Oct 1938.

W. G. Cochran. Matching in Analytical Studies. *American Journal of Public Health*, 43, Jun 1953.

W. G. Cochran. The Planning of Observational Studies of Human Populations. *Journal of the Royal Statistical Society, Ser. A*, 1965.

G. Colak and T. M. Whited. Spin-offs, Divestitures, and Conglomerate Investment. *The Society for Financial Studies*, 20(3), 2006.

M. J. Cooper, H. Gulen, and P. R. Rau. Changing Names with Style: Mutual Fund Name Changes and Their Effects on Fund Flows. *The Journal of Finance*, 60(6): pp. 2825–2858, Dec 2005.

H. F. Dorn. Philosophy of Inferences from Retrospective Studies. *American Journal of Public Health*, 43, Jun 1953.

S. Drucker and M. Puri. On the Benefits of Concurrent Lending and Underwriting. *The Journal of Finance*, 60(6):pp. 2763–2799, Dec 2005.

E. F. Fama and K. R. French. Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics*, 33(1):3 – 56, 1993.

A. Gande and M. Puri. Enhancing Security Value by Ownership Restrictions: Evidence from a Natural Experiment. *Financial Management*, 34(4):pp. 35–64, Winter 2005.

G. Gong, H. Louis, and A. X. Sun. Earnings Management and Firm Performance following Open-Market Repurchases. *The Journal of Finance*, 63(2):pp. 947–986, Apr 2008.

R. G. Greenberg. The Use of Analysis of Covariance and Balancing in Analytical Surveys. *American Journal of Public Health*, 43, Jun 1953.

E. Greenwood. *Experimental Sociology: A Study in Method*. King's Crown Press, 1945.

G. Grullon and D. L. Ikenberry. What do we know about Stock Repurchases? *Journal of Applied Corporate Finance*, 13(1):31–51, 2000.

G. Grullon and R. Michaely. Dividends, Share Repurchases, and the Substitution Hypothesis. *The Journal of Finance*, 57(4):1649–1684, 2002.

J. Heckman and S. Navarro-Lozano. Using Matching, Instrumental Variables, and Control Functions to Estimate Economic Choice Models. *The Review of Economics and Statistics*, 86, 2004.

T. Hellmann, L. Lindsey, and M. Puri. Building Relationships Early: Banks in Venture Capital. *The Review of Financial Studies*, 21(2):pp. 513–541, Mar 2008.

C. E. Hogan and C. M. Lewis. Long-Run Investment Decisions, Operating Performance, and Shareholder Value Creation of Firms Adopting Compensation Plans Based on Economic Profits. *The Journal of Financial and Quantitative Analysis*, 40(4):pp. 721–745, Dec 2005.

D. Ikenberry, J. Lakonishok, and T. Vermaelen. Market Underreaction to Open Market Share Repurchases. *Journal of Financial Economics*, 39(2-3):181 – 208, 1995.

V. Ivashina, V. B. Nair, A. Saunders, N. Massoud, and R. Stover. Bank Debt and Corporate Governance. *The Review of Financial Studies*, 22(1):pp. 41–77, Jan 2009.

N. Jegadeesh and J. Karceski. Long-run Perormance Evaluation: Correlation and Heteroskedasticity-Consistent Tests. *Journal of Empirical Finance*, 16(1), Jan 2009.

M. C. Jensen. Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers. *The American Economic Review*, 76(2):pp. 323–329, May 1986.

S. P. Kothari and J. B. Warner. Measuring Long-Horizon Security Price Performance. *Journal of Financial Economics*, 43(3), Mar 1997.

C. M. Kuhnen. Business Networks, Corporate Governance, and Contracting in the Mutual Fund Industry. *The Journal of Finance*, 64(5):pp. 2185–2220, Oct 2009.

X. Li and X. Zhao. Propensity Score Matching and Abnormal Performance After Seasoned Equity Offerings. *Journal of Empirical Finance*, 13(3), Jun 2006.

J. D. Lyon and B. M. Barber. Detecting Long-Run Abnormal Stock Returns: The Empirical Power and Specification of Test Statistics. *Journal of Financial Economics*, 43, 1997b.

J. D. Lyon, B. M. Barber, and C. Tsai. Improved Methods for Tests of Long-Run Abnormal Stock Returns. *The Journal of Finance*, 54(1), 1999.

A. Marosi and N. Massoud. "You Can Enter but You Cannot Leave...": U.S. Securities Markets and Foreign Firms. *The Journal of Finance*, 63(5):pp. 2477–2506, Oct 2008.

R. W. Masulis, C. Wang, and F. Xie. Agency Problems at Dual-Class Companies. *The Journal of Finance*, 64(4):pp. 1697–1727, Aug 2009.

W. J. McNally and B. F. Smith. Long-Run Returns following Open Market Share Repurchases. *Journal of Banking & Finance*, 31(3):703 – 717, 2007.

M. H. Miller and F. Modigliani. Dividend Policy, Growth, and the Valuation of Shares. *The Journal of Business*, 34(4):pp. 411–433, Oct 1961.

M. H. Miller and K. Rock. Dividend Policy under Asymmetric Information. *The Journal of Finance*, 40(4):pp. 1031–1051, Sep 1985.

C. C. Peters. A Method of Matching Groups for Experiment with No Loss of Population. *Journal of Educational Research*, 34(8), 1941.

U. Peyer and T. Vermaelen. The Nature and Persistence of Buyback Anomalies. *Review of Financial Studies*, 22(4):1693–1745, 2009.

P. R. Rosenbaum and D. B. Rubin. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, 70(1), Apr 1983.

D. B. Rubin. The Use of Matched Sampling and Regression Adjustment to Remove Bias in Observational Studies. *Biometrics*, 29, Mar 1973.

D. B. Rubin. Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies. *The Journal of the American Statistical Association*, 74(366), Jun 1979.

B. Villagonga. Does Diversification Cause the "Diversification Discount"? *Financial Management*, 33(2), 2004.

VITA

Sunayan Acharya

Date and place of birth

06 February 1982, Calcutta, India

Educational institutions attended and degrees awarded

August 2007 - present

University of Kentucky, Ph.D. in Finance

August 2000 - June 2004

Indian Institute of Technology Madras, B.Tech in Electrical Engineering)

Professional positions held

Teaching and Research Assistant, Department of Finance, University of Kentucky, 2008 - present

Teaching and Research Assistant, Department of Physics and Astronomy, University of Kentucky, 2004 - 2007