The Pennsylvania State University
The Graduate School

ESSAYS IN FINANCIAL ECONOMICS

A Dissertation in
Business Administration
by
Zhijian Huang

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

August 2008

The dissertation of Zhijian Huang was reviewed and approved* by the following:

Jingzhi (Jay) Huang
David H. McKinley Professor of Business Administration and Finance
Dissertation Advisor, Co-Chair of Committee

Jeremy Ko
Assistant Professor of Finance
Co-Chair of Committee

Jean Helwege
Associate Professor of Finance

Runze Li
Associate Professor of Statistics

William J. Kracaw
David Sykes Professor of Finance
Chair, Department of Finance

*Signatures are on file in the Graduate School.

Abstract

Previous studies show mixed results about the out-of-sample performance of various asset-pricing anomalies. To reduce data-snooping bias, this paper simulates a real-time trader who chooses among all asset-pricing anomalies published prior to that time using only non-forward-looking filters. I find that a trader can outperform the market by recursively picking the best past performer among published anomalies. The excess return tends to be highest when the trader looks at past performances between two years and five years and when the trader considers more anomalies. For published anomalies, their excess returns over benchmark as well as relative ranks among contemporaneous anomalies do not decrease over time, indicating a relatively stable performance once being published. Relying only on the then-available anomaly literature and historical data, the overall result shows a possible way to beat the market in real time.

In the second essay, I study the managerial decision and its impact on mutual fund performance. "Meet or beat" describes manager's manipulation of the accounting earnings to meet analyst forecasts, which causes the final earnings distribution to be negatively-skewed with a dent to the left. I investigate if mutual fund managers also demonstrate such a risk-seeking behavior and the potential

impact on mutual fund performance. I develop a two-period decision model in which a manager makes a decision to take or not to take a risky project when she faces a target level of performance in time 1. The manager's payoff depends on a linear function of the time 2 (final) value of the firm, and on a step function on whether the firm meets the target or not. As a result, those managers with firm values slightly lower than expected tend to take the risky project, and if the firm value is already higher or far below the expectation, the manager does not take the project. The terminal value of a firm conforms with the empirical evidence of meeting or beating analyst forecasts. I empirically test the model on actively managed mutual funds. I find funds in the 3rd and 4th worst performance deciles take higher risks and on average outperform the median decile in the following year, indicating a risk seeking behavior for slightly under-performing managers. I also compare the excess return distributions for growth funds and index funds, but do not find negative skewness supporting the model.

The third essay is an experimental study about people's time-inconsistent risk preferences. There is a substantial literature which studies time-inconsistent temporal preferences. We conducted an experiment to explore time-inconsistency in the other dimension of investment preferences, i.e., risk preferences. We had subjects play a multi-period betting game where they planned their betting decisions in advance and then played the game dynamically later to see if these decisions matched their plan. We found that subjects took more risk than planned in their initial bet and after a loss where this increase in risk-taking is associated with an increase in breakeven mental accounting. Our findings shed light on the conditions under which emotions exacerbate mental accounting and other behavioral biases.

Table of Contents

List of Figures

List of Tables

Acknowledgments

First, my thanks go to my dissertation committee members for their guidance. My thesis advisor Jingzhi (Jay) Huang not only gave me advice, but also kept track of all the details of my research. He also taught me how to communicate effectively and how to survive in this profession. His encouragement has enabled me to overcome many crises over the past several years. I am grateful to Jeremy Ko, who gave me support by generously providing his time, advice, and guidance throughout my academic career. Jeremy Ko has spent a lot of his time reading different versions of my working papers. I learn from him how to write clearly, precisely, and concisely. I also thank Jean Helwege whose helpful suggestions and assistance greatly improved my research, particularly in the monthly progress meetings she conducted with Jingzhi (Jay) Huang, which helped me to form my own research agenda. Special thanks go to my outside committee member Runze Li who provides insights from a different perspective.

I also want to thank other faculty members in the finance department at Penn State University. Laura Field, as the Ph.D. coordinator, supported me during the most difficult time of my Ph.D. years. Olesya Grishchenko, David Haushalter, Hong Li, Tim Simin, John Wald, and Vivian Wang gave me useful comments to improve my research as well as my presentation skills.

Toni Auman, Christina Jones, and Jennifer Sunderland provided me a lot of help on many administrative issues.

My wife, Liping Jia, accompanied me all the way through this journey. Without her, I would not have been able to finish this dissertation. I also want to acknowledge all my friends and family members who gave me support through my years of study.

Chapter 1

Introduction

One interesting feature in finance research is the finding of asset-pricing anomalies, empirical facts that could not be fully explained by rational models. Explanations of market inefficiency mainly fall into two categories: rational and behavioral approaches. Rational explanations attribute the existence of anomalies to the difficulty rational arbitrageurs face in trading away price discrepancies. In contrast, behavioral explanations simply conclude that market participants do not behave rationally as opposed to the assumptions of most financial models. This dissertation contributes in both directions.

The first essay, which is on the rational side, addresses the data-snooping bias in the out-of-sample test of asset-pricing anomalies. Previous studies show mixed results about the out-of-sample performance of various asset-pricing anomalies. To reduce data-snooping bias, this paper simulates a real-time trader who chooses among all asset-pricing anomalies published prior to that time using only non-forward-looking filters. I find that a trader can outperform the market by recursively picking the best past performer among published anomalies. The excess return tends to be highest when the trader looks at past performances between two years and five years and when the trader considers more anomalies. For published anomalies, their excess returns over benchmark as well as relative ranks among contemporaneous anomalies do not decrease over time, indicating a relatively stable performance once being published. Relying only on the then-available anomaly literature and historical data, the overall result shows a possible way to beat the market in real time.

In the second essay, I give a rational explanation for the suboptimal managerial decision when a manager faces a target level of performance. "Meet or beat" describes manager's manipulation of the accounting earnings to meet analyst forecasts, which causes the final earnings distribution to be negatively-skewed with a dent to the left. I investigate if mutual fund managers also demonstrate such a risk-seeking behavior and the potential impact on mutual fund performance. I develop a two-period decision model in which a manager makes a decision to take or not to take a risky project when she faces a target level of performance in time 1. The manager's payoff depends on a linear function of the time 2 (final) value of the firm, and on a step function on whether the firm meets the target or not. As a result, those managers with firm values slightly lower than expected tend to take the risky project, and if the firm value is already higher or far below the expectation, the manager does not take the project. The terminal value of a firm conforms with the empirical evidence of meeting or beating analyst forecasts. I empirically test the model on actively managed mutual funds. I find funds in the 3rd and 4th worst performance deciles take higher risks and on average outperform the median decile in the following year, indicating a risk seeking behavior for slightly under-performing managers. I also compare the excess return distributions for growth funds and index funds, but do not find negative skewness supporting the model.

The third essay turns to the behavioral side by exploring people's time-inconsistent risk preferences in a laboratory experiment. There is a substantial literature which studies time-inconsistent temporal preferences. We conducted an experiment to explore time-inconsistency in the other dimension of investment preferences, i.e., risk preferences. We had subjects play a multi-period betting game where they planned their betting decisions in advance and then played the game dynamically later to see if these decisions matched their plan. We found that subjects took more risk than planned in their initial bet and after a loss where this increase in risk-taking is associated with an increase in breakeven mental accounting. Our findings shed light on the conditions under which emotions exacerbate mental accounting and other behavioral biases.

Overall, the findings in this dissertation show that while rational and behavioral reasons can account for some of the anomalies in financial markets, but not all.

Chapter 2

Real-Time Profitability of Published Anomalies: An Out-of-Sample Test

2.1 Introduction

In the study of asset-pricing anomalies, data snooping may occur when researchers draw inferences from the single realization of a historical time series (e.g., see Lo and MacKinlay, 1990). One way to guard against data snooping bias is to conduct out-of-sample tests.[1] However, Cooper and Gulen (2006) point out that even out-of-sample tests are not totally immune from the data snooping bias. In fact, they show that given the large freedom in the research design of out-of-sample tests, a researcher may unintentionally report time-series predictability that is sensitive to the exogenously selected simulation parameters. This performance sensitivity to research design is also reflected in the mixed empirical findings in the real-time out-of-sample test literature. Breen, Glosten, and Jagannathan (1989), Brennan and Xia (2001), Cooper (1999), Cremers (2002), Solnik (1993), and Xia (2001) find that dynamic asset allocation in real time can beat the market. On the other hand,

[1] Other methods to assess the likelihood of a data snooping bias include applying statistical model selection criteria (Bossaerts and Hillion, 1999) and considering the whole universe from which a model is drawn (White, 2000).

other authors including Cooper, Gutierrez, and Marcum (2005), Goyal and Welch (2003, 2007), Handa and Tiwari (2006), and Pesaran and Timmermann (1995) find the real-time profitability inconclusive. While the research designs of these papers differ from each other in many aspects, one important factor affecting the out-of-sample performance is the choice of predictive variables under study.[2] Data snooping could arise when a researcher chooses which variables to study based on foresight or knowledge of the entire time series. This paper tries to remove this potential data-snooping bias by simulating a real-time trader who searches the best trading strategies using all published anomalies prior to an evaluation time. I seek to address the question of whether a rational trader who efficiently processes information in real time can realize excess returns based *only* on asset-pricing anomalies published prior to the evaluation time.

In this paper, I run a rolling window out-of-sample test in which I assume a real-time trader chooses trading strategies from published anomalies only. In other words, the simulated trader takes an agnostic view on which predictive variables to use and relies on the sophisticated procedure of academic research to unfold anomalies over time. Specifically, I simulate a trader progressively picks the best performing anomaly over time out of an expanding universe of anomalies as they become published. In this simulation, the anomaly universe is determined in real time by publications in academic journals after applying some non-forward-looking filters. Specifically, I limit my search within anomalies that 1) are published in several finance journals that are top ranked in real time, 2) are calendar anomalies or about cross-sectional predictability, and 3) can be re-evaluated annually.[3] Therefore, only those anomalies that have already been published enter into the real-time trader's scope. For example, the strategy of

[2]Cooper and Gulen (2006) summarize in a "reality spectrum" three major choice variables in an out-of-sample test. In addition to the predictive variables under test, the other two important variables are the benchmark to compare with and the in-sample training length.

[3]More details about these filters will be provided in Section 2.2, where I also give examples to show why these filters select anomalies differently than in other papers.

"investing in small firms" is not considered until 1981, when Banz (1981) first publishes his seminal work about the size anomaly. This process guarantees that any excess returns a simulated trader earned in the past did not benefit from foresight. By considering appropriate transaction costs, I then compare the real-time profit of chasing the best published anomaly to a benchmark return.

My major empirical finding is that a trader who trades with the best published anomalies based on backtesting can earn considerable returns in excess of a market benchmark. More specifically, I find that in different cases, a real-time trader can significantly beat the market in annual excess return by average margins ranging from 4.43% to 13.23%. This result is based on annual model re-selection among anomalies that were available at the time. Among different training lengths, the excess return is higher if the trader back tests with a training period of two to five years. This superior performance also comes with better statistics in other commonly used performance measures such as the one-factor alpha, the Sharpe ratio, and the certainty equivalent rate of returns. Similar results are obtained when the trader considers only a certain subgroup of anomalies and with different levels of transaction cost. Also, as I increase the number of anomalies available in the real-time trader's total anomaly universe, I see a steady increase in the average of realized excess returns. Therefore, considering more published asset-pricing anomalies seems to be a good practice. As to the characteristics of anomalies being chosen in real time, I find "old" anomalies published several years ago perform just as well as "new" anomalies that have just been published, in terms of both absolute and relative measures.

This paper contributes to the literature by proposing a novel methodology to reduce data snooping bias in out-of-sample tests. In this simulation, real-time decisions such as which predictive variables to consider (i.e., to back test) and to trade with are completely based on the past information relative to an evaluation point. In the existing real-time investment literature, some papers test too few variables, leaving a large possibility of data snooping. To show different results about the real-time profitability, a researcher can choose or not choose a particular anomaly based on foresight, that is, knowledge of which anomalies persist or do not persist. For example, Brennan and Xia (2001) and Goyal and Welch (2003) reach opposite conclusions about the real-time profitability by each studying just one asset-pricing anomaly: the Fama-French three factor model for the former and the return predictability of dividend yield for the latter. In contrast, some other studies endogenize the search for best model specification into a too large of a search space by enumerating all possible combinations of predictive variables. For example, Cooper, Gutierrez, and Marcum (2005) and Pesaran and Timmermann (1995) each assess 170 and 511 different model specifications, respectively.[4] One problem of this type of specification search is the lack of intelligent reduction in the search space of model specifications. It is very hard to explain economically why a complex combination of predicative variables would be the best out-of-sample predictor for the next period. As Cooper and Gulen (2006) point out, with a huge number of different combinations of key parameters, even out-of-sample predictability could be the result of data snooping after a massive (and inadvertent) specification search by researchers.

Another contribution of this paper is about the persistence of published anomalies. Schwert (2003) finds that while some asset-pricing anomalies have persisted after initial publication, most of them have attenuated or disappeared either as a result of data snooping or because of the widespread awareness of these anomalies. The question remains as to the overall verdict these results pose for market efficiency. I find that in spite of this attenuation, a trader can indeed earn abnormal profits by back testing publicly known asset-pricing anomalies in real time.

In this paper, even though I remove the data-snooping bias arising from the

[4] There are other papers which search an even larger space. For example, Cremers (2002) has a combinatorial model space with $2^{14} = 16,384$ models.

choice of anomalies, there is still some freedom left in the simulation design that may potentially cause data snooping. I try to control the remaining data-snooping bias in two ways. First, throughout this research, I report or at least test all scenarios associated with each particular choice of exogenous variable. For example, I report results with different training lengths, different transaction costs, and different subsets of the anomaly universe. Moreover, rather than highlighting the best result of the simulations in this study, I focus on the most "straight forward" cases. For example, I assume the real-time trader considers primarily the one-year training period unless otherwise convinced. I allow for the real-time trader to pick a suboptimal parameter relative to what we can find in hindsight. Second, I test the robustness of my main conclusions using different sample periods. I switch the calendar back to an earlier time point and repeat the whole set of studies. By doing this, I can see whether or not I would have found the same results had I conducted this research earlier.

The idea of only using the then-available technology in out-of-sample tests has been mentioned in several real-time investment papers. For example, Cooper, Gutierrez, and Marcum (2005) add the once believed Beta strategy into trader's strategy universe to make the simulation more realistic even though the Beta strategy no longer works. Pesaran and Timmermann (1995) briefly discuss that a trader can only use the data as well as the technology available at a historical point in time. They also argue that the macroeconomic variables studied in their paper should have been known to investors since 1950's and early 1960's. However, the nine basic predictive variables used in their paper are still chosen "based on a review of the early literature" (Pesaran and Timmermann, 1995, page 1208). To the best of my knowledge, no paper has imposed the restriction that a real-time trader could notice an anomaly *if and only if* after it has been published.[5] This

---

[5]In fact, many anomalies were known to the public long before their academic publications. Here the publication requirement is simply a backward-looking filter to reduce the data-snooping bias in selecting anomalies. In other words, for an existing anomaly, one can interpret academic publication as a proxy for reaching a certain level of publicity after a careful scrutiny by the

paper differs from the out-of-sample test literature by endogenizing the choice of models to only published anomalies from a naive anomaly universe based on non-forward-looking filters.

The rest of the paper starts with a description of the research design and the in-sample performances in Section 2.2. Section 2.3 reports the out-of-sample results along three lines: the length of training period, the transaction cost, and the number of anomalies under consideration. Section 2.4 analyzes which anomalies are chosen more frequently relative to others in real time. Based on two new measures developed to reflect the competitiveness of anomalies among contemporaneous anomalies, I also investigate if there is any time trend in anomaly performances after publication. Section 2.5 conducts some robustness checks including an "out-of-sample test" on this research and tests assuming traders learn about anomalies slightly earlier or later than the publication years. Section 2.6 concludes the research and discusses some issues regarding the position of this paper in empirical finance research.

2.2 Research Design

Just as Cooper and Gulen (2006) pointed out, the traditional way of picking anomalies to study, which is primarily based on their importance in the recent finance literature, may cause data snooping bias. I correct this problem by mimicking a real-time trader's anomaly selection process that looks at academic publications before an evaluation time. In each year, the trader searches through a predefined subset of finance journals for implementable trading strategies. All asset-pricing anomalies published before a historical time point draw equal attention to the trader. Based on his own back testing results of all published anomalies, the trader determines which anomaly to be used for the next calendar

academic research procedure.

year.

To reduce the number of anomalies to a feasible range without incurring forward-looking bias, I apply three filters to the anomaly selection process: the anomaly type, the publishing journal, and the rebalancing frequency.[6] These three filters are not forward-looking in the sense that they do not rely on future performance or future publicity in selecting anomalies. First, I limit my choice of anomalies to those within two broad categories of asset-pricing anomalies: calendar anomalies and cross-sectional return anomalies. These anomalies are relatively easy to back test because they only use market return and firm characteristics data that was widely available when these anomalies were first published.

Second, I assume the real-time trader in my simulation only looks for anomalies in five major finance journals: the Journal of Finance (JF), the Journal of Financial Economics (JFE), the Review of Financial Studies (RFS), the Journal of Financial and Quantitative Analysis (JFQA), and the Journal of Business (JB).[7] I then look through every paper in the above finance journals during the time between 1972 and 2005 searching for anomalies that might have caught the attention of a real-time trader. I choose 2005 as the end year for searching simply because the data is available through 2006. 1972 is chosen as the starting point for two reasons. First, the last seminal paper formulating the foundation of CAPM was published by Black (1972), after which the real-time trader could have a clear definition of an anomaly. Second, 1972 marks the first empirical test on CAPM, which

[6]An implicit filter in this research is the implementability of anomalies. Some anomalies cannot be exploited without leverage or short selling. For example, Penman (1987) reports higher returns for stocks during the first two weeks of April, July, and October. Kim and Park (1994) find abnormally high stock returns before holidays. However, these stock return seasonalities cannot be taken advantage of by a trading strategy without borrowing in margin.

[7]While I cannot cover all journals to mimic the true opportunity set of a real-time trader, I did try to expand this five-journal list by including the references in the papers chosen from the above four journals, regardless of where those references were originally published. This is to assume that if a paper in one of the four journals is chosen, the real-time trader reads it carefully enough to include all referenced papers. This alternative assumption only results a slight difference in the availability of anomalies so that one anomaly (the book-to-market ratio predictor) is known to the real-time trader much earlier. The real-time trader's performance under this assumption is therefore slightly stronger.

verifies the linear relationship between portfolio returns and their betas (Black, Jensen, and Scholes, 1972). I rely on this issue-by-issue search to find the true set of anomaly universe faced by a real-time trader over time. This search process ensures that anomalies are chosen because they caught a trader's attention in real time, regardless whether or not they are still famous nowadays. Finally, to reduce the small sample bias, my third filter excludes those long-term return anomalies that require a holding period of more than one year to generate excess returns.

Choosing anomalies based on academic publications generates three major differences relative to the traditional way of selecting anomalies. First, small journal publications which become famous later should not be considered in real time, because the trader does not know which papers later would be famous and which would not. An example is the earliest reference about the book-to-market ratio predictability. Stattman (1980) in journal, "The Chicago MBA: A Journal of Selected Papers", first reports the book-to-market ratio anomaly. However, unlike academics who benefit in hindsight from the growing literature on asset-pricing anomalies, the simulated trader (in this study with a limited vision of only five journals) would not know the anomaly until it was first mentioned by Chan, Hamao, and Lakonishok (1991) in the Journal of Finance. The second difference is that top journal publications that are no longer famous in the anomaly literature will be considered in real time. One example is the debt/equity ratio predictability by Bhandari (1988, JF). Although it is not widely believed that leverage ratio can predict equity returns, the real-time trader would seriously consider this anomaly because it is published in a top finance journal. Finally, an anomaly becomes known at the time of its first publication, not the time of its most famous publication, if both papers are in top journals. For example, the momentum strategy would be known in 1990 by the publication of Jegadeesh (1990, JF), rather than in 1993 by Jegadeesh and Titman (1993, JF), although the latter is more famous and receives more citations.

The following two subsections introduce the basic anomalies studied in this research in detail and provide a description of the recursive out-of-sample simulation. I focus on several key elements of this research design: training and holding periods, anomaly selection criteria, transaction costs, and performance measures.

2.2.1 Data and the Basic Anomalies

There are 11 basic anomalies implemented in this research. Two of them are calendar anomalies: the Monday effect and the January effect. The Monday effect, also known as the weekend effect, says stock returns are on average lower on Mondays relative to other week days. Therefore, a corresponding trading strategy could be to hold T-Bills on Monday and to hold a market index on other weekdays. The January effect, also called the turn-of-the-year effect, says stock returns are higher in January, especially for small firms. One implementable strategy could be buying small firms in January and holding a market index for the rest of the year. The rest of the nine basic anomalies implemented in this study are cross-sectional stock return anomalies based on seven predictive variables: size, book-to-market ratio (B/M), momentum, earnings/price ratio (E/P), cashflow/price ratio (CF/P), dividend yield (Div/P), debt/equity ratio (D/E), growth in sales (GS), and trading volume (measured by average turnover ratios). To implement a cross-sectional strategy, a trader simply needs to hold a specific risk factor decile of stocks rather than the market index during the trading period.

I use annual value-weighted returns of the NYSE/AMEX/NASDAQ index from the Center for Research in Security Prices (CRSP) as the market benchmark return. The risk-free rates are the one month T-Bill rates also from CRSP. For the returns on cross-sectional portfolios, I use the monthly factor risk portfolio returns data from Kenneth French's data library[8] for six out of the nine predictive

[8]I acknowledge Kenneth French for making these data sets available at his Web site at:

variables: size, book-to-market (B/M) ratio, momentum, earnings/price ratio (E/P), cashflow/price ratio (CF/P), and dividend yield (Div/P). The factor risk portfolio returns based on the debt/equity ratio, the growth in sales, and the trading volume are not available in Kenneth French's data library. Therefore, I form decile portfolios and calculate their returns from the CRSP/Compustat database. Below are the details about the calculation of these values:

- The Debt/Equity ratio

  The D/E ratio is calculated following Bhandari (1988):

$$\text{D/E ratio} = \frac{\text{book value of total assets - book value of common equity}}{\text{market value of common equity}} \tag{2.1}$$

  The book value of total assets is Compustat data item 6. The book value of common equity is calculated following the procedure in Cohen, Polk, and Vuolteenaho (2002). The market value of common equity is the year end share price multiplied by the number of shares outstanding in CRSP. Following the standard of the cross-sectional return literature (e.g., Fama and French (1993)), I assume the accounting variables become available six months after the end of the fiscal year. In other words, the accounting variables for fiscal year $t-1$ are used to form portfolios from July of the year t to June of the year $t+1$. Also, I only use the NYSE stocks to calculate the break points for deciles.

- The Growth in Sales

  Following Lakonishok, Shleifer, and Vishny (1994, LSV94 here after), I calculate the growth in sales by weighting the sales growth ranks for the past five years. Specifically, I extract the total sales of a firm from the Compustat data item 12 to calculate the annual growth rates for the past five years. All

---

http://mba.tuck.dartmouth.edu/ pages/faculty/ken.french/data_library.html.

the growth rates in each year are then ranked in relative percentage from 0% to 100%. To calculate the weighted growth in sales, I multiply the last year growth rank by 5, the growth rank two years ago by 4, the growth rank three years ago by 3, etc, and sum them together. All the stocks in each year are divided into deciles according to their growth in sales. According to the finding in LSV94, those firms with low sales growth rate are value stocks with a higher expected return in future years. Similarly, I set break points using only NYSE stocks and assume the newly calculated accounting variable becomes available in the next July after each year.

- The Trading Volume

  The trading volume is measured by the average turnover ratio over a year as in Lee and Swaminathan (2000). Following their approach, I assume the real-time trader considers only NYSE and AMEX stocks. Unlike those accounting variables such as the D/E ratio and the growth in sales, trading volume information is immediately available at the end of each calendar year. Therefore, portfolio rebalancing occurs in January of each year. According to Lee and Swaminathan (2000), low volume stocks have higher future expected returns.

For the market returns and the risk free rate, the data range is from 1926 to 2006; the factor risk portfolios have different starting years, but all end in 2006. Table 2.1 lists the initial publication of each anomaly, the data availability range, and a brief description of the trading strategy implemented for each basic anomaly.

Table 2.1. Information about the Basic Anomalies

This table summarizes the basic information about the 11 anomalies implemented in this research by listing their initial publications, data availability, and brief descriptions of the associated trading strategy. JFE is for the Journal of Financial Economics and JF is for the Journal of Finance. Panel A lists two strategies based on calendar anomalies. Panel B lists nine trading strategies based on cross-sectional predictive variables: size, book-to-market (B/M) ratio, lagged return (short-term momentum and long-term reversal), earnings/price ration (E/P), cash flow/price ratio (CF/P), dividend yield (Div/P), debt/equity ratio (D/E), the growth in sales (GS), and trading volume measured by the average turnover ratio.

	Panel A: Strategies Based on Calendar Anomalies		
Anomaly Name	Initial Publication	Data Range	Strategy
Monday Effect	French (1980, JFE)	1926-2006	Hold T-Bills on Monday; hold the market index on other weekdays
January Effect	Keim (1983, JFE)*	1926-2006	Hold smallest decile stocks for January; hold the market for the rest of the year

	Panel B: Strategies Based on Cross-Sectional Return Predictability		
Anomaly Name	Initial Publication	Data Range	Strategy**
Size	Banz (1981, JFE)	1926-2006	Buy small firms
B/M	Chan, Hamao, and Lakonishok (1991, JF)	1926-2006	Buy firms with high B/M ratio
Momentum	Jegadeesh (1990, JF)	1926-2006	Buy winners in last year
E/P	Basu (1977, JF)	1952-2006	Buy firms with high E/P ratio
CF/P	Fama (1990, JF)	1952-2006	Buy firms with high cash flow
Div/P	Ball (1978, JFE)	1928-2006	Buy firms with high dividend yield
D/E	Bhandari (1988, JF)	1965-2006	Buy firms with high debt/equity ratio
GS	Lakonishok, Shleifer, and Vishny (1994, JF)	1970-2006	Buy firms with low sales growth rate
Volume	Lee and Swaminathan (2000, JF)	1964-2006	Buy firms with low average turnover ratio

* The anomaly implemented here is also called the “turn-of-the-year” effect. Since I categorize it as a calendar anomaly, I choose the name “January effect” but there is a difference from the anomaly in Rozeff and Kinney (1976).

** In brief descriptions for cross-sectional return strategies, “buy” means holding a decile portfolio of a particular risk factor throughout the trading period.

Table 2.2. The In-Sample Performance of Basic Anomalies

This table demonstrates the in-sample performance of each basic anomaly. *Publication Year* is the year of the first publication for that anomaly. *Excess Before* (*Excess After*) is the average annual excess return over the benchmark return before (after) publication. *Terminal Wealth* is the trader's terminal wealth in dollar amount at the end of year 2006 if he starts with \$1 in the publication year after considering transaction costs. *Wealth Bench* is the terminal wealth for the benchmark index. *Sharpe Ratio* and *Sharpe Bench* are sharpe ratios for the trading strategy and the benchmark index after publication, respectively. The Sharpe ratio is calculated by deviding the mean excess return over risk free rate by its standard deviation: $Sharpe = \frac{mean(R_p - R_f)}{Std(R_p - R_f)}$ where R_p is the annual return of the strategy and R_f is the annual risk free rate.

Anomaly	Publication Year	Excess Before	Excess After	Terminal Wealth †	Wealth Bench	Sharpe Ratio	Sharpe Bench
Monday	1980	9.42%***	-0.65%	5.49	25.54	0.48	0.51
JanEffect	1983	13.02%***	9.82%***	87.8	16.23	1.2	0.55
Size	1981	11.46%***	1.51%	15.59	18.92	0.36	0.48
B/M	1991	5.29%**	6.19%	10.95	5.21	0.78	0.56
Momentum	1990	9.23%***	7.92%**	12.45	4.83	0.62	0.47
E/P	1977	10.71%***	6.12%**	124.28	33.78	0.78	0.49
CF/P	1990	6.71%***	3.73%*	7.85	4.83	0.67	0.47
Div/P	1978	2.89%*	1.4%	38.18	34.84	0.6	0.53
D/E	1988	8.16%**	6.31%*	18.76	7.81	0.72	0.54
GS	1994	6.04%**	1.43%	3.41	2.82	0.8	0.47
Volume	2000	7.09%***	16.4%***	2.28	0.17	1.19	0.04

***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

† In calculating the terminal wealth, a 5% annual transaction cost is charged to the Monday effect and 2% for the January effect. All the rest strategies have a 1% annual transaction cost.

As we can see from Table 2.1, most anomalies were published in the 1970s and the 1980s. Note that the publications listed in the second column may not be the earliest publication of the anomaly. Since one filter used in the anomaly selection procedure is to only include anomalies in four top finance journals, there is a possibility that an anomaly may have appeared in another journal at an earlier time.

Table 2.2 lists the in-sample performances of the basic anomalies before and after their initial publication. The anomalous return is measured by the annual average excess return over a benchmark index. We can see all anomalies experienced a decrease in anomalous returns after being published except for the volume predictor which has only six years of sample after its publication in 2000. For example, when compared with the benchmark index after its publication year, the Monday effect has a negative excess return.[9] For other anomalies, they deliver reduced or less significant excess returns after initial publication. These in-sample results are generally consistent with the 2003 survey article by Schwert (2003), which summarizes the reduced performances of asset-pricing anomalies after their publication. Therefore, it is a non-trivial question to test whether or not a real-time trader can outperform the market using only trading strategies based on published anomalies.

2.2.2 The Rolling Window Trading Scheme

In my simulation, I follow the real-time investment literature to implement a rolling window model selection and trading scheme. In addition, I set the initial publication of an anomaly as the time point when the anomaly becomes publicly available.[10] Below is a narrative description about the design of the recursive

[9] Actually, there seems to have been a "regime switch" for the Monday effect in 1989 after which Monday performs the best, rather than the worst, among the weekdays. See Rubinstein (2001, page 25).

[10] In Section 2.5, I perform robustness checks with anomalies available to public slightly earlier or later than their publications.

trading scheme, with key elements highlighted and discussed in details later.

First, I form an anomaly universe using part or all of the 11 basic anomalies. The simulation starts at the earliest publication year within the anomaly universe, that is, a trader started to trade when at least one anomaly has been published. The trader picks an anomaly if it is the only one available (i.e., it is the only anomaly ever published.). If there is more than one anomaly, the trader picks the one that performs the best over a past *training period* according to a particular *strategy selection criterion*.[11] After selecting the best anomaly, the trader trades according to that anomaly for the next *holding period*. At the end of the holding period, the trader liquidates his portfolio and repeats the above procedure standing at a time point one *holding period* later with an enlarged information set. The simulation ends at year 2006, which is the latest year with available data. After the simulation, the performance of the real-time trader is then compared with the performance of a market benchmark return using different *performance measures*. During the trading process, different levels of *transaction cost* are applied to different trading strategies, but there is no transaction cost considered for the buy-and-hold strategy on benchmark portfolio. Figure 2.1 illustrates an example of this out-of-sample test scheme with a 10-year training period and a one-year holding period.

- Training and Holding Periods

  I study the rolling training (or portfolio formation) period of the past one year, two years, five years, and ten years. Although not formally listed, results based on other training lengths are also tested and used to generate figures in this paper. For the holding period, I only test the one year case.

[11]Here I assume the trader believes in active trading so that even if none of the anomalies outperform the market during the training period, he will still pick the one with the least underperformance rather than indexing for the next year. Actually, the real-time trader may find that no anomaly works in the past one year standing at the beginning of 1980, 1991, and 2006. As a robustness check, I release this assumption and find the performances are slightly worse if the trader ever decides to simply index in some years.

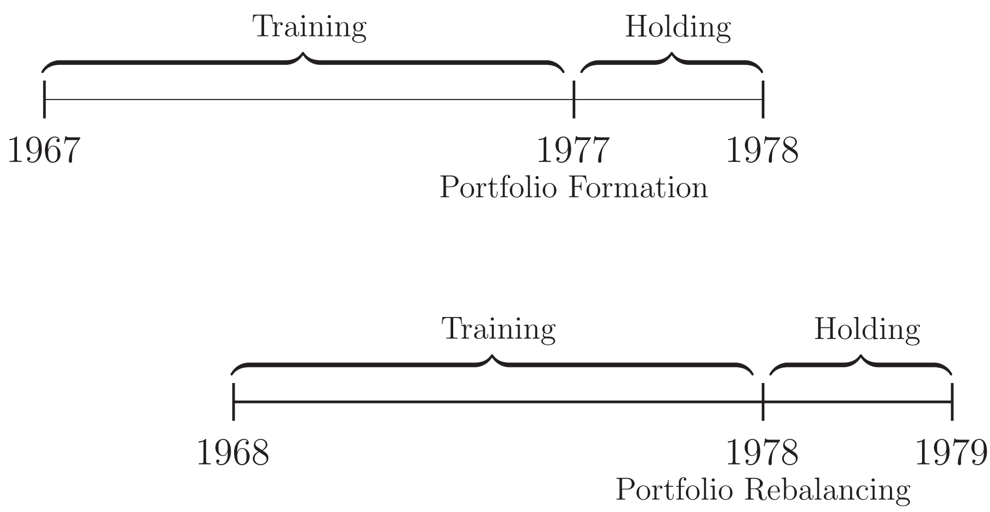

Figure 2.1. An illustration of the rolling window out-of-sample test from year 1977 to 1978. In this example, the training period is 10 years and the holding period is one year. Portfolios are formed every year based on the backtesting results of all published anomalies by that time.

The reason is that in this simulation the trader starts trading only after those anomalies being published in the 1970s or 1980s, leaving only about 30 years for the simulation. Therefore, it is efficient to assume a one year holding period and annual portfolio rebalancing to fully utilize the available data.

- Strategy Selection Criteria

  In this research, I use the past excess return as the primary strategy selection criterion. In the robustness check Section 2.5, I also test the Sharpe ratio criterion as is used in, for example, Cooper, Gutierrez, and Marcum (2005) and Pesaran and Timmermann (1995). I do not use any statistical criteria such as those in Bossaerts and Hillion (1999) for model selection. To be consistent with my assumption that the real-time trader learns about anomalies based on academic publications, I assume the trader treats each anomaly as a "black box" without really making judgments based on its structural characteristics. This assumption eliminates the possibility that the trader will evaluate the fitness of an anomaly based on its internal structure, such as the number of predictive parameters in the model. Instead, the real-time trader relies purely on the real-time back testing results to make investment decisions.

- Transaction Costs

  Transaction cost in implementing trading strategies is a very important component in out-of-sample tests. However, it is hard to estimate precisely the transaction costs for various anomalies over time. To address this issue, first I set up a base-level transaction cost for each anomaly. Then I test different transaction cost levels by adjusting the base transaction costs of various anomalies.

  Following Pesaran and Timmermann (1995), I assume a 1% transaction cost on a round-trip transaction between risk-free T-Bills and stock shares,[12] and a 2% transaction cost between different stocks. This is the base level of transaction cost. As we know from Table 2.1, except for the Monday effect, all the basic strategies in this study have an annual turnover ratio of two (i.e., one round-trip transaction). Also, among those strategies with two annual turnover, all but the January effect incur transactions between the risk-free asset and stocks. Therefore, the transaction cost is 2% annually for the January effect and 1% for the rest strategies except for the Monday effect.

  The Monday effect has such a high turnover rate that even with the most conservative estimate of transaction cost, it is not a practical trading strategy. Nonetheless, the Monday effect is such an important calendar anomaly that

[12] Precisely it is 1.1% in Pesaran and Timmermann (1995) because they also impose a 0.1% transaction cost on T-Bills.

a real world investor can hardly neglect it. To solve this dilemma, I assume a five percent base level transaction cost for the Monday effect. One possible explanation could be that only institutional investors could implement this strategy at a highly discounted transaction cost due to both the economy of scale and the use of derivatives.

In addition to the base level, I test three transaction cost variations. First, I try different values for the cost of round-trip transactions between stocks and T-Bills. Specifically, I assume there is 0%, 0.5%, 1.5%, or 2% one-way turnover cost in addition to the 1% base level. Second, I customize transaction costs across different trading strategies. In this paper, there are three anomalies incur transactions with small or illiquid stocks: the size anomaly, the January effect, and the volume anomaly. Trading with small stocks requires a higher transaction cost as documented by, for example, Stoll and Whaley (1983) and Bhardwaj and Brooks (1992). Therefore, I set the round-trip transaction cost of these three anomalies to 5%. Finally, I apply the real-world transaction cost data from Stoll (1993) and French (2008). These transaction costs are estimated by dividing the aggregated revenue of all brokerage and dealer firms by the total value of securities traded in each year. According to French (2008, Table V), The one-way transaction cost decreases over time from 146 basis points in 1980 to only 11 basis points in 2006. I extrapolate the data linearly for the years 1977 to 1979.

- Performance Measures

To measure the real-time trader's performance relative to the market index, I formally report four performance measures. The first performance measure is the mean annual excess return over the benchmark return. It measures how much, on average, the trader can beat the market in real time. The second measure is the annual *alpha* from the one-factor market model:

$$R_p - R_f = \alpha + \beta(R_m - R_f) + \epsilon \tag{2.2}$$

where R_p, R_f, and R_m are returns from the portfolio of the real-time trader, the risk-free return, and the market returns, respectively.[13]

The third measure is the Sharpe ratio difference between the simulated return and the market return, measuring the risk-return characteristics of the real-time trader:

$$\begin{aligned} \text{Diff Sharpe} &= Sharpe_p - Sharpe_m \\ &= \frac{mean(R_p - R_f)}{Std(R_p - R_f)} - \frac{mean(R_m - R_f)}{Std(R_m - R_f)} \end{aligned} \tag{2.3}$$

Finally, I report the certainty equivalent rate (CER) of return used in, for example, Handa and Tiwari (2006). It is a risk adjusted return for a particular group of risk averse traders. Again, we measure the CER differences between the real-time trader's portfolio and the benchmark portfolio as:

$$\begin{aligned} \text{Diff CER} &= CER_p - CER_m \\ &= [mean(R_p) - mean(R_m)] - \frac{1}{2}\alpha[Var(R_p) - Var(R_m)] \end{aligned} \tag{2.4}$$

Here α is the risk aversion level of the trader. In this study, I set $\alpha = 2$.

I calculate statistical significance based on a standard t test for the first two measures: the annual excess return and the one-factor Alpha. For the other two measures, I calculate the significance numerically by bootstrapping.[14]

[13] In this paper I do not consider the alpha in multi-factor models, because an "anomaly" in this research is defined as any abnormal return not theoretically explainable by the CAPM model. Multi-factor models are empirical asset-pricing models emerged from the abnormal portfolio returns based on size, book-to-market ratio, and momentum. In this paper, these factors are supposed to be exploited in real time rather than serving as performance benchmarks.

[14] In fact, there is a closed form solution for the asymptotic distribution of the Sharpe ratio

I only calculate an empirical p-value together with a sample probability density distribution based on multiple simulations with different subsets of the basic anomalies.

- Short Sale and Borrowing on Margin

  In this study, I design the trading strategies so as not to allow for short sales and borrowing on margin. The primary concern is that a lot of the real world market participants are restricted from leverage and short selling, such as mutual funds. However, not allowing for short selling could raise the suspicion of data snooping because many anomalies, such as the momentum strategy, were originally proposed to form a "zero-investment" portfolio by opening both long and short positions. Nonetheless, given the major empirical finding of this research that a trader could beat the market in real time, not allowing any leverage or short selling is in fact a stronger assumption than allowing them, because short-selling constraints reduce portfolio performances as illustrated in Alexander (2000).

2.3 Out-of-Sample Results

I present the out-of-sample performance of the real-time trader in three aspects. The first part is the out-of-sample performance of the real-time trader with different training lengths. In the second subsection, I report the performance with varying transaction costs to see how sensitively the out-of-sample result varies to the cost of trading. Finally, to see the relationship between excess return and the number of

difference. If we assume the excess returns (over the risk-free rate) are i.i.d., the Sharpe ratio difference converges to a distribution of $N(\frac{\mu_p}{\sigma_p} - \frac{\mu_m}{\sigma_m}, \frac{2}{N})$, where μ_p and σ_p represent the true mean and standard deviation respectively of the excess return for the real-time trader, and μ_m and σ_m are the mean and standard deviation for the market. Under the null hypothesis that $\mu_p = \mu_m$ and $\sigma_p = \sigma_m$, the statistical test is simply a z-test with mean zero and the standard deviation of $\sqrt{2/N}$, where N is the number of years in the simulation. The numerical results yield almost the same significance as the theoretical ones.

anomalies considered, I run out-of-sample simulations with only a random subset of anomalies being included. Unless stated otherwise, the transaction costs used in these simulations are based on a 1% round-trip transaction cost as described in Section 2.2.2.

2.3.1 Performance with Different Training Lengths

Table 2.3 lists the simulation result of a trader who picks an available (i.e., published) anomaly based on past performances with various training lengths. I report the real-time trader's returns and the benchmark returns for three different anomaly groups: all anomalies, calendar anomalies only, and cross-sectional return anomalies only. I list four performance measures relative to the benchmark performance: the excess return over the benchmark return, the one-factor alpha, the difference in Sharpe ratios, and the difference in certainty equivalent rate (CER) of returns. Statistical significance for the excess return and the one-factor alpha measures is based on a standard *t*-test. The significance for the differences in Sharpe ratio and certainty equivalent return is estimated numerically using bootstrapping.

As we can see, in all scenarios in Table 2.3 the real-time trader who chases the performance of published anomalies beats the market as indicated by the positive values of all four measures in all cases. The average annual excess return over the buy-and-hold return of a market benchmark ranges from 4.43% to 13.23% and is significant at the 10% confidence level under all scenarios. If the real-time trader happened to choose two years or five years as the training length, the performance would be even better—with excess returns significant at the 1% level and Alpha's significant at the 5% level for these six cases.[15] The simulation from year 1977 (1980 for the calendar anomalies only case) also generates a larger Sharpe ratio

[15] The positive Alpha does not come with increased risk. The one-factor Beta ranges from 0.82 to 1.20 in all 12 cases.

Table 2.3. The Out-Of-Sample Performance with Different Training Lengths

This table presents the out-of-sample performance of a trader who picks the best performing anomaly among published anomalies recursively over time. Panels A through D represent different training lengths of 1, 2, 5, and 10 years, respectively. I report three different anomaly groups represented by *Group Name*: *All* is for all anomalies; *Calendar* is for the two calendar anomalies only; *Cross-Sec* is for the nine strategies based on cross-sectional return predictability. For each anomaly group, I report the average annual return (*Raw Return*) and the average annual benchmark return (*Return Bench*). I also report four performance measures of the real-time return over the benchmark index: the annual excess return (*Excess Return*), the one-factor *Alpha*, the difference in Sharpe ratios between the real-time trader and the market index (*Diff Sharpe*), and the difference of the certainty equivalent rate of returns between the real-time trader and the market index with a risk aversion level of 2 (*Diff CER*). The detailed calculations of *Diff Sharpe* and *Diff CER* are described in Section 2.2.2. The standard deviations are listed in brackets.

Group Name	Raw Return	Bench Return	Excess Return	Alpha	Diff Sharpe	Diff CER
	Panel A: 1-Year Training Period					
All	18.06%	13.63%	4.43%**	4.66%**	0.17	3.72%
	(17.62%)	(15.47%)	(10.35%)	(2.15%)	(0.3)	(4.42%)
Calendar	20.24%	14.04%	6.21%**	8.1%***	0.35	6.1%*
	(16.16%)	(15.83%)	(11.82%)	(2.49%)	(0.34)	(4.54%)
Cross-Sec	18.72%	13.63%	5.08%**	4.5%*	0.13	3.63%
	(19.63%)	(15.47%)	(12.09%)	(2.5%)	(0.31)	(4.73%)
	Panel B: 2-Year Training Period					
All	24.49%	13.63%	10.86%***	9.66%***	0.27	7.73%*
	(23.49%)	(15.47%)	(16.71%)	(3.43%)	(0.29)	(5.01%)
Calendar	21.5%	14.04%	7.47%***	9.93%***	0.44*	7.45%*
	(15.91%)	(15.83%)	(13.06%)	(2.69%)	(0.36)	(4.5%)
Cross-Sec	23.07%	13.63%	9.44%***	8.35%**	0.22	6.51%*
	(23.07%)	(15.47%)	(16.4%)	(3.37%)	(0.28)	(4.93%)
	Panel C: 5-Year Training Period					
All	26.86%	13.63%	13.23%***	13.64%***	0.47**	11.35%***
	(20.67%)	(15.47%)	(16.15%)	(3.35%)	(0.3)	(4.53%)
Calendar	22.58%	14.04%	8.54%***	10.79%***	0.52*	8.58%**
	(15.71%)	(15.83%)	(12.19%)	(2.52%)	(0.37)	(4.54%)
Cross-Sec	24.46%	13.63%	10.83%***	10.54%***	0.34	8.66%**
	(21.37%)	(15.47%)	(15.51%)	(3.22%)	(0.28)	(4.45%)
	Panel D: 10-Year Training Period					
All	22.48%	13.63%	8.85%***	10.15%***	0.48*	8.61%**
	(16.24%)	(15.47%)	(11.53%)	(2.33%)	(0.33)	(4.25%)
Calendar	21.22%	14.04%	7.18%***	8.9%***	0.43*	7.2%*
	(15.77%)	(15.83%)	(10.78%)	(2.27%)	(0.36)	(4.49%)
Cross-Sec	18.73%	13.63%	5.1%*	4.52%	0.09	3.13%
	(20.88%)	(15.47%)	(14.13%)	(2.92%)	(0.31)	(4.98%)

***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

and a higher CER return than the market benchmark in all cases.

Figure 2.2 shows the terminal wealth evolving over time for four different training lengths. Starting from \$1 in 1977, in 2006 a trader could have reached from \$100 to more than \$800 by actively back testing and trading with published anomalies only, net of transaction cost. In contrast, indexing in a broad market portfolio will only generate \$37.11. The 2-year and 5-year training lengths seem to do a better job identifying profitable anomalies in real time. However, the length of the training period does not seem to have a monotonic relation with out-of-sample profitability. Figure 2.3 shows details about the effect of training length.

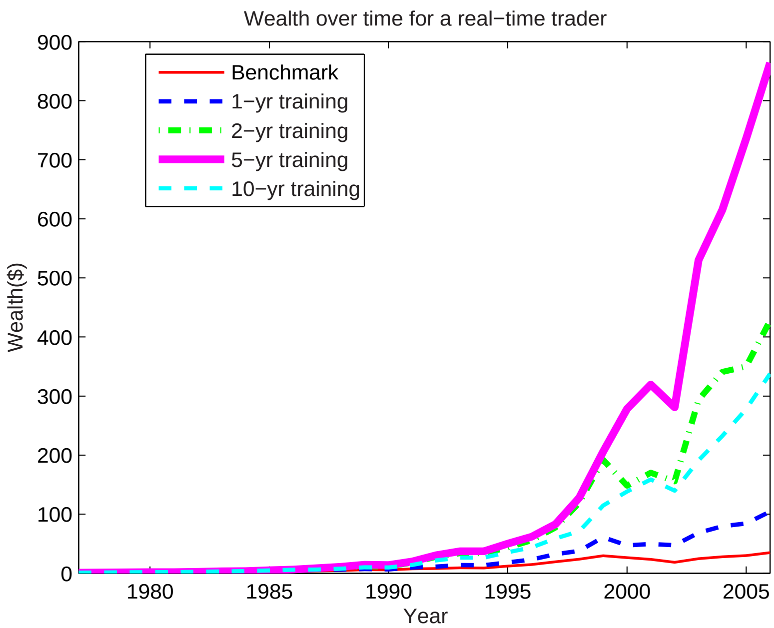

Figure 2.2. Terminal wealth of the market index and for a real-time trader who starts with \$1. The training lengths are 1 year, 2 years, 5 years, and 10 years.

By aggregating the results in Table 2.3 plus some unreported cases using other training lengths, I report the relationship between training length and excess return in Figure 2.3. The three lines correspond to all anomalies, calendar anomalies only, and cross-sectional return anomalies only. As shown in Figure 2.3, neither a too short (1 year) nor a too long (more than 5 years) training length is optimal for the

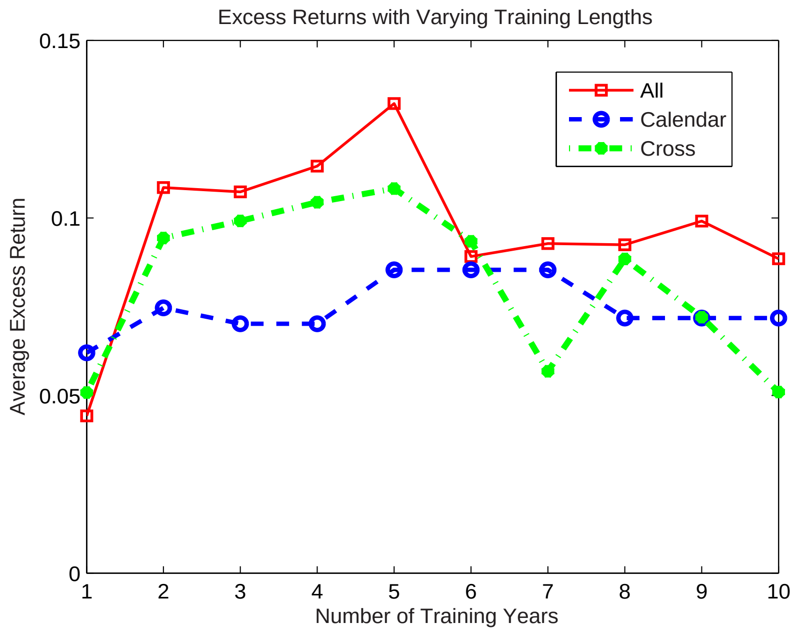

Figure 2.3. Excess returns over the benchmark index for different training lengths. The X axis represents different training lengths from 1 year to 10 years. The Y axis is the excess return over the benchmark index.

real-time trader if he picks from all anomalies or from cross-sectional anomalies only. In this study, a training period between two years and five years seems to deliver better out-of-sample performances.

2.3.2 Performance under Different Transaction Costs

The finding that a real-time trader can beat the market by only considering published anomalies is based on several assumptions outlined in Section 2.2.2. One of them is the 1% transaction cost on a one-way turnover between stocks and T-Bills, which is the "high level" transaction cost in Pesaran and Timmermann (1995). In this section I test the performance sensitivity to different schemes of transaction cost. Table 2.4 presents the performances with all anomalies considered under different transaction cost levels. Different panels represent various training lengths.

There are five fixed transaction cost levels in Table 2.4 ranging from 0% to 2%, including the 1% level used in other parts of the paper. I test the customized transaction cost level in which a 5% transaction cost is imposed on the size anomaly, the January effect, and the volume effect, because these strategies incur transactions with small and illiquid stocks. The last row of each panel applies the real world annual transaction cost data from French (2008) which is estimated from the aggregated stock brokerage revenue and the total trading volume.

As expected, the real-time trader's performances are eroded by an increasing transaction cost in Table 2.4. However, the main conclusion of this paper—a positive and significant excess return and alpha—persists in most of the cases, including the case using real transaction cost data. Actually, the increased transaction cost barely affect the significance of the results if the trader has a training length longer than one year. Unreported tests show that the case with five-year training length will remain significantly better than the market benchmark even the round-trip transaction cost rises to as high as 5% for all anomalies. The break even threshold transaction costs for the two-year and ten-year training lengths to remain statistically significant at the 10% level are 3% and 3.5%, respectively.

2.3.3 Performance with a Subset of Basic Anomalies

As mentioned in the introduction, the way to remove data-snooping bias in this research is to closely mimic the true anomaly universe perceived by a trader in real time. However, I have to add several filters to reduce the number of anomalies to an implementable level. Although the three filters used in this research are designed not to be forward-looking, it is still an interesting question to see what if the real-time trader knows fewer anomalies than the 11 anomalies implemented in this research. The reason for such an exercise is that each trader, in reality, may only focus on a limited number of anomalies based on his expertise, rather

Table 2.4. Out-Of-Sample Performances under Different Transaction Costs

This table presents the out-of-sample performance of a trader who picks the best past performer among the 11 basic anomalies recursively over time. Panels A through D represent different training lengths of 1, 2, 5, and 10 years, respectively. I report cases with seven different schemes of transaction cost. The first five rows in each panel use a fixed round-trip transaction cost of 0%, 0.5%, 1%, 1.5%, and 2%, respectively. The case *Customized* assumes the same base level transaction cost for all strategies except that it is 5% for the size anomaly, the January effect, and the volume effect. The *real data* case uses the annual estimated transaction costs from French (2008). For each level of transaction cost, I report the average annual return (*Raw Return*) and the average annual benchmark return (*Return Bench*). I also report four performance measures of the real-time return over the benchmark index: the annual excess return (*Excess Return*), the one-factor *Alpha*, the difference in Sharpe ratios between the real-time trader and the market index (*Diff Sharpe*), and the difference of the certainty equivalent rate of returns between the real-time trader and the market index with a risk aversion level of 2 (*Diff CER*). The detailed calculations of *Diff Sharpe* and *Diff CER* are described in Section 2.2.2.

Transaction Costs	Raw Return	Bench Return	Excess Return	Alpha	Diff Sharpe	Diff CER
	Panel A: 1-Year Training Period					
0%	19.53%	13.63%	5.9%***	6.19%***	0.26	5.24%
0.5%	18.79%	13.63%	5.16%***	5.42%**	0.21	4.48%
1%	18.06%	13.63%	4.43%**	4.66%**	0.17	3.72%
1.5%	17.33%	13.63%	3.7%*	3.89%*	0.12	2.95%
2%	16.59%	13.63%	2.96%	3.12%	0.08	2.17%
Customized	16.93%	13.63%	3.3%*	3.64%*	0.12	2.67%
Real Data	18.22%	13.63%	4.59%**	4.89%**	0.19	3.92%
	Panel B: 2-Year Training Period					
0%	26.36%	13.63%	12.72%***	11.48%***	0.34	9.58%**
0.5%	25.42%	13.63%	11.79%***	10.57%***	0.31	8.66%**
1%	24.49%	13.63%	10.86%***	9.66%***	0.27	7.73%*
1.5%	23.56%	13.63%	9.92%***	8.75%**	0.23	6.79%*
2%	22.62%	13.63%	8.99%***	7.85%**	0.19	5.84%
Customized	23.22%	13.63%	9.59%***	8.27%**	0.21	6.39%
Real Data	25.05%	13.63%	11.42%***	10.17%***	0.29	8.3%*
	Panel C: 5-Year Training Period					
0%	28.46%	13.63%	14.83%***	15.28%***	0.54**	12.95%***
0.5%	27.66%	13.63%	14.03%***	14.46%***	0.5**	12.15%***
1%	26.86%	13.63%	13.23%***	13.64%***	0.47*	11.35%***
1.5%	26.06%	13.63%	12.43%***	12.81%***	0.43*	10.54%***
2%	25.26%	13.63%	11.63%***	11.99%***	0.39*	9.73%**
Customized	24.52%	13.63%	10.89%***	11.36%***	0.35	8.89%**
Real Data	27.15%	13.63%	13.52%***	13.98%***	0.49**	11.7%***
	Panel D: 10-Year Training Period					
0%	24.22%	13.63%	10.58%***	11.87%***	0.57**	10.34%***
0.5%	23.35%	13.63%	9.72%***	11.01%***	0.52**	9.47%**
1%	22.48%	13.63%	8.85%***	10.15%***	0.48*	8.61%**
1.5%	21.62%	13.63%	7.98%***	9.29%***	0.43*	7.74%**
2%	20.75%	13.63%	7.12%***	8.43%***	0.37*	6.87%**
Customized	20.02%	13.63%	6.38%***	7.61%***	0.32	6.04%*
Real Data	22.91%	13.63%	9.28%***	10.56%***	0.51**	9.04%**

***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

than all anomalies published in academic journals. To address this issue, I shrink the anomaly universe to a random subset of the 11 basic anomalies in this section. Two questions can be asked: First, can the real-time trader still beat the market if not all anomalies are considered? Second, is considering more anomalies always good?

I repeat the simulation in Section 2.3.1 for cases when the trader only knows a subset of the original 11-anomaly universe. This gives us a total of $2^{11} - 1 = 2047$ different simulations for the all-anomaly case, 3 simulations for the case with calendar anomalies only, and 511 simulations for the case with cross-sectional anomalies only. Table 6 reports the performances of these simulations sorted by the number of anomalies considered in the reduced anomaly universe. Based on different performance measures, the real-time trader outperforms the market even when only a subset of anomalies are ever considered.

As we can see, for the all-anomaly case, the excess return measure and the difference in certainty equivalent rate of return measure generate better performance when more anomalies are considered. For the case with only cross-sectional anomalies, we can see similar results. However, we do not see a monotonic trend in the one-factor Alpha and the Sharpe ratio measures. Instead, there is a "U-shaped" performance for the real-time trader. The simulations yield a high Alpha and a high Sharpe ratio difference if very few anomalies are considered or almost all anomalies are considered. Good performance with more anomalies is consistent with the other two measures. This "U-shaped" performance indicates that as the average excess return increases with the number of anomalies considered, so does the return variance. The return variance increases faster than the excess return when the number of anomalies is small. However, when the number of anomalies is large, the return variance increases less rapidly.

The fundamental research question of this paper is: Can a trader beat the market in real time by following published anomalies? In Section 2.3.1 we find

Table 2.5. The Out-Of-Sample Performance with a Subset of Basic Anomalies

This table presents the out-of-sample performances of a trader who picks the best performing anomaly out of random subsets of the some basic anomalies recursively over time. Panel A studies the group with all anomalies. Panel B and Panel C represent groups with calendar anomalies only and cross-sectional anomalies only, respectively. The training period for this table is one year. Within each panel, I run simulations of having only a subset of the total available anomalies. The number of selected anomalies, *Number of Anomalies*, ranges from one anomaly to the total number of anomalies available within that group. *Number of Combi.* is the total number of different combinations of anomalies to form a subset of the group. I report the average annual return *Average Raw Rt*, the average annual benchmark return *Average Bench Rt*. I also report four performance measures averaging across all combinations. *Excess Return* is the average annual excess return. *Average Alpha* is the average one-factor Alpha. *Avg Diff Sharpe* is the difference in Sharpe ratios between the real-time trader and the market index. *Avg Diff CER* is the difference of the certainty equivalent rate of returns between the real-time trader and the market index with a risk aversion level of 2. The detailed calculations of *Diff Sharpe* and *Diff CER* are described in Section 2.2.2.

Number of Anomalies	Number of Combi.	Average Raw Rt	Average Bench Rt	Excess Return	Average Alpha	Avg Diff Sharpe	Avg Diff CER
Panel A: All Anomalies							
1	11	16.42%	12.4%	4.02%	5.54%	0.19	3.38%
2	55	17.07%	13.53%	3.54%	4.92%	0.14	2.62%
3	165	17.42%	13.7%	3.72%	4.8%	0.13	2.68%
4	330	17.63%	13.79%	3.85%	4.66%	0.12	2.72%
5	462	17.77%	13.82%	3.95%	4.53%	0.11	2.77%
6	462	17.86%	13.82%	4.04%	4.41%	0.11	2.83%
7	330	17.92%	13.8%	4.11%	4.32%	0.11	2.91%
8	165	17.96%	13.77%	4.18%	4.29%	0.12	3.03%
9	55	17.99%	13.73%	4.26%	4.32%	0.13	3.2%
10	11	18.02%	13.68%	4.34%	4.45%	0.15	3.43%
11	1	18.06%	13.63%	4.43%	4.66%	0.17	3.72%
Panel B: Calendar Anomalies							
1	2	14.96%	13.88%	1.08%	3.26%	0.08	1.06%
2	1	20.24%	14.04%	6.21%	8.1%	0.35	6.1%
Panel C: Cross-Sectional Anomalies							
1	9	16.74%	12.07%	4.67%	6.04%	0.22	3.89%
2	36	16.79%	13.34%	3.44%	4.64%	0.12	2.3%
3	84	16.95%	13.54%	3.41%	4.25%	0.09	2.09%
4	126	17.13%	13.67%	3.46%	4.03%	0.08	2.04%
5	126	17.33%	13.73%	3.6%	3.91%	0.08	2.13%
6	84	17.58%	13.75%	3.83%	3.9%	0.08	2.34%
7	36	17.89%	13.73%	4.15%	4%	0.09	2.67%
8	9	18.27%	13.69%	4.57%	4.19%	0.11	3.11%
9	1	18.72%	13.63%	5.08%	4.5%	0.13	3.63%

Note: All values reported to the right of the vertical line are the average values over all cases represented by the *Number of Combinations*.

positive answers in terms of several performance measures. This section, in addition, provides another angle to evaluate the statistical significance of the major finding of this paper. Basically, we can assess the significance based on different choices about anomaly subsets. I aggregate the 2047 simulation results corresponding to all subsets of the 11 basic anomalies to form empirical probability density functions (PDFs) of the four performance measures: the excess return over benchmark, the one-factor Alpha, the Sharpe difference, and the certainty equivalent rate of return difference. Based on the empirical PDFs, I calculate the empirical p-values for the four performance measures. Figure 2.4 lists the PDFs and their corresponding p-values. As we can see, the real-time trader who happens to know a random subset of the 11 basic anomalies can almost always generate excess return relative to the benchmark as well as a positive Alpha in a one-factor model. If measured by Sharpe ratio or the CER rate of returns, the performances with different subsets are slightly worse but still gives the real-time trader a big chance to beat the market.[16]

2.4 Anomaly Choices

The previous section shows that even without data snooping, a real-time trader can still outperform the market using only published asset-pricing anomalies. This section further analyzes this result by looking into the anomaly choices made by the trader in the process of real-time back testing and trading. Specifically, two questions are investigated.

First, I want to discover which anomalies are more frequently chosen. Since anomalies become available in different publication years, the widely used "inclusion frequency" measure (e.g., see Pesaran and Timmermann (1995) and Cremers

[16]Another performance measure, the terminal wealth in excess of the market benchmark, generates a similar picture with an empirical p-values of 0.0352. For the convenience of presentation, here I do not show it graphically.

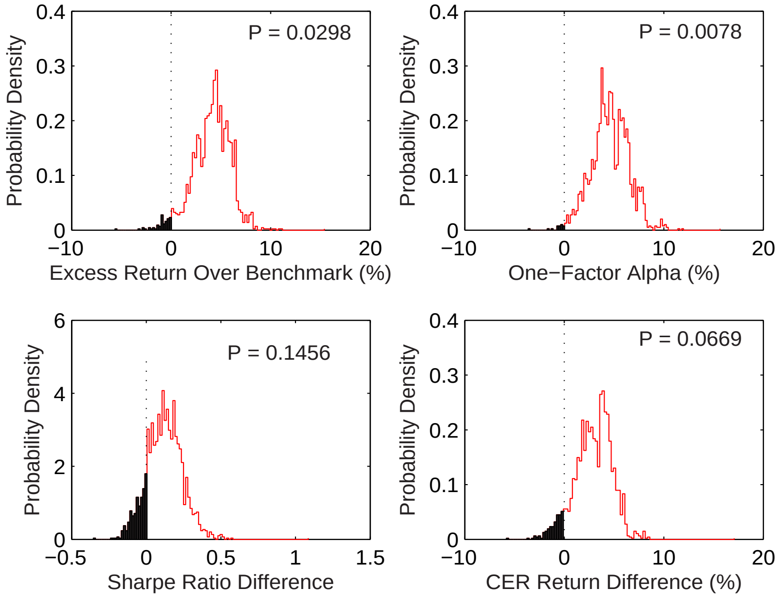

Figure 2.4. Sample probability density functions (PDFs) for four performance measures: the excess return, the one-factor Alpha, the Sharpe ratio difference, and the CER return difference. These PDFs are estimated using simulations based on the 2047 subsets of the 11 basic anomalies. Empirical p-values are reported based on the ratio of negative values among all simulation runs.

(2002)) cannot truly reflect the relative effectiveness of each anomaly in real time. Here I develop the "relative popularity" (RP) measures which reflect an anomaly's performance rank relative to other contemporaneous anomalies. Also, I look at the most "popular" anomalies to see if they are the same ones we find in-sample and if those anomalies dominate the overall results.

Second, using the relative popularity measures, I investigate whether or not "old" anomalies published several years ago are less likely to be chosen compared to "new" anomalies that have just been published. In other words, I am interested in if there is any time trend in performance for anomalies after publication, in terms of either absolute profitability measured by excess return or performance relative to other anomalies. If asset-pricing anomalies decrease in profitability after publication over time, a simulated real-time trader, having successfully exploited published anomalies in the past, may find them unprofitable in the future. Schwert (2003) focuses on the in-sample performance to show that asset-pricing anomalies generally disappear or attenuate after their initial publication. The question addressed here is therefore whether this post-publication performance drop happens immediately or gradually over time.

The following subsections present the detail about the mathematical definition of relative popularity measures and empirical results, organized by the above two areas of inquiry.

2.4.1 Which Anomalies are More Frequently Chosen?

Fixing other parameters, a proper measure of the competitiveness of an anomaly in the scope of a real-time trader should satisfy the following three properties: 1) it increases as the number of years being chosen increases; 2) it decreases as the anomaly's total number of years available increases; and 3) it increases as the number of other anomalies in the chosen years increases. I develop the relative popularity of an anomaly based on its normalized rank among all published anomalies. Over a period of time, the measure can be roughly described as:

$$\text{Relative Popularity}(t_1, t_2) = \frac{\text{number of anomalies outperformed}}{\text{number of anomalies that could have outperformed}} \tag{2.5}$$

Specifically, I calculate two types of relative popularity. The first type (RP1) only counts first place winners in each year. The second type (RP2), however, assigns points to all published anomalies based on their ranks even when the anomaly is not chosen. For example, in 1981, only four anomalies had been published and therefore were available to the real-time trader. Ranked from the best to worst by their performances in the previous year (i.e., in 1980), these anomalies were: the Monday effect, the size anomaly, the E/P ratio anomaly, and the dividend yield anomaly. A real-time trader using a one-year training length would find the E/P ratio anomaly ranked third among the four anomalies published in or before 1981. According to RP1, the E/P ratio anomaly gets zero points for 1981 because it is not the best anomaly in that year, but it will get some points using RP2 because the E/P ratio anomaly still outperforms the dividend yield anomaly in rank for the year 1981. Below are the detailed definitions of these two values for an anomaly i in the period between years t_1 and t_2, using a training period of j years:

$$RP1_{i,j}(t_1, t_2) = \frac{\sum_{\tau=t_1}^{t_2} (N^\tau - 1) \cdot I(r_{i,j}^\tau = 1) \cdot (N^\tau/2)}{\sum_{\tau=t_1}^{t_2} (N^\tau - 1)} \tag{2.6}$$

$$RP2_{i,j}(t_1, t_2) = \frac{\sum_{\tau=t_1}^{t_2} (N^\tau - r_{i,j}^\tau)}{\sum_{\tau=t_1}^{t_2} (N^\tau - 1)} \tag{2.7}$$

when $\sum_{\tau=t_1}^{t_2} (N^\tau - 1) > 0$, that is, there is more than one anomaly in at least one year. Here N^t is the number of anomalies available in year t, $I(\cdot)$ is the indicator function, and $r_{i,j}^t$ is the real-time perceived rank of anomaly i, in year t, based on a j-year training length, and in which better anomalies have lower rankings. Both RP1 and RP2 equal to their unconditional mean of 1/2 if only one anomaly ever existed the period of time between years t_1 and t_2. Mathematically, if $\sum_{\tau=t_1}^{t_2} (N^\tau - 1) = 0$, $RP1_{i,j}(t_1, t_2) = RP2_{i,j}(t_1, t_2) = 1/2$. For the relative popularity in a particular year, the above notations can be simplified as:

$$RP1_{i,j}^t \equiv RP1_{i,j}(t, t) = I(r_{i,j}^t = 1) \cdot (N^t/2) \tag{2.8}$$

$$RP2_{i,j}^t \equiv RP2_{i,j}(t, t) = \begin{cases} \frac{N^t - r_{i,j}^t}{N^t - 1} & \forall \ N^t > 1 \\ \frac{1}{2} & \text{if} \ N^t = 1 \end{cases} . \tag{2.9}$$

Note that there is a normalization factor $N^t/2$ in the definition of $RP1_{i,j}^t$ so that given a particular year t, both RP1 and RP2 have an average value of 1/2 across all published anomalies in that year, because when $N^t > 1$ for any t:

$$\overline{RP1}^{\,t} \equiv \frac{1}{N^t}\sum_{i=1}^{N^t} RP1_{i,j}^t = \frac{1}{N^t} I(r_{i,j}^t = 1)\cdot(N^t/2) = \frac{1}{2} \tag{2.10}$$

$$\overline{RP2}^{\,t} \equiv \frac{1}{N^t}\sum_{i=1}^{N^t} RP2_{i,j}^t = \frac{1}{N^t}\sum_{i=1}^{N^t}\frac{N^t - r_{i,j}^t}{N^t - 1} = \frac{1}{N^t}\sum_{i=1}^{N^t}\frac{N^t - i}{N^t - 1} = \frac{1}{2}. \tag{2.11}$$

Therefore, the expected values of both relative popularity measures are irrelevant to time t. If the rankings of published anomalies are random, there should not be any time trend in the relative popularity measures.

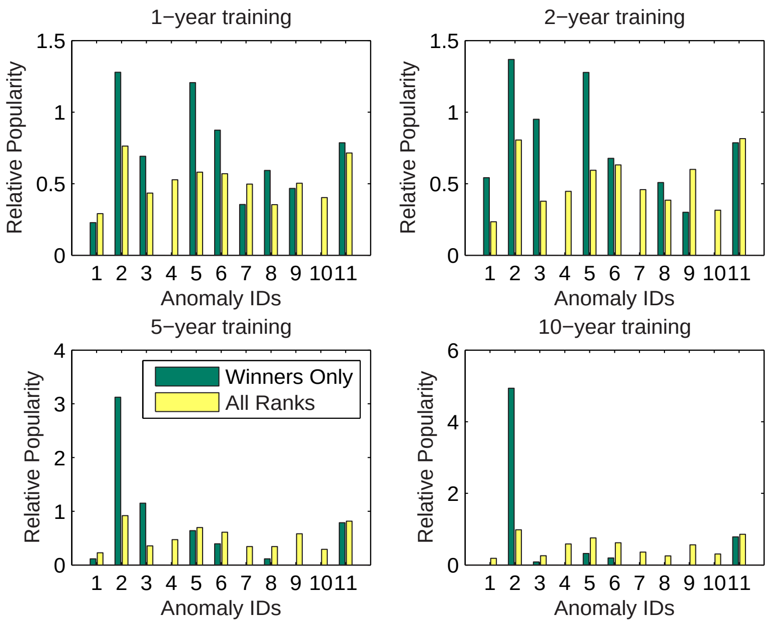

Figure 2.5. The life time relative popularity of each anomaly when a trader picks from all published anomalies in real time. The Y-axis is the number of years an anomaly is used. The X-axis represents different anomalies indexed by a sequence number: 1 for the Monday effect, 2 for the January effect, 3 for the size variable, 4 for the B/M variable, 5 for the momentum variable, 6 for the E/P variable, 7 for the CF/P variable, 8 for the Div/P variable, 9 for the Debt/Equity variable, 10 for the growth in sales variable, and 11 for the volume variable.

Figures 2.5 and 2.6 present the life time relative popularity of anomalies using four different training lengths: one year, two years, five years, and ten years.

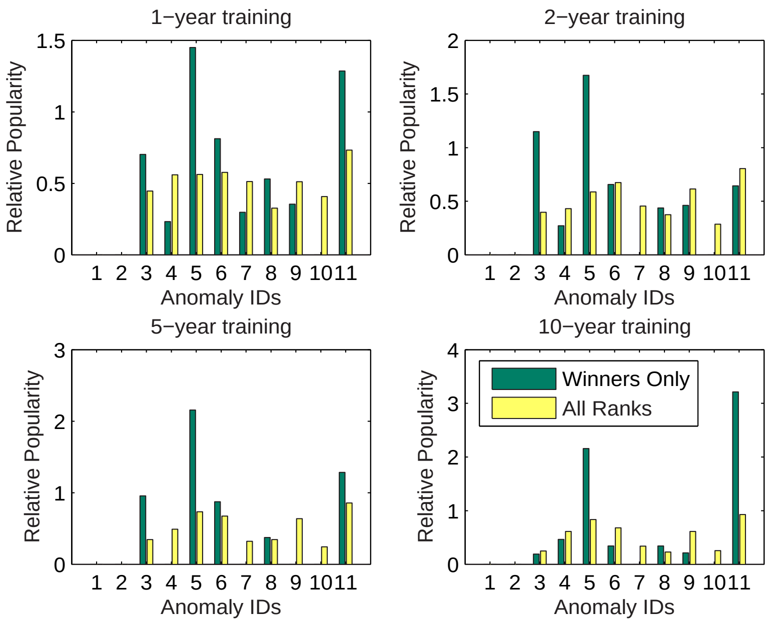

Figure 2.6. The life time relative popularity of each anomaly when a trader picks from only the cross-sectional anomalies in real time. The Y-axis is the number of years an anomaly is used. The X-axis represents different anomalies indexed by a sequence number: 3 for the size variable, 4 for the B/M variable, 5 for the momentum variable, 6 for the E/P variable, 7 for the CF/P variable, 8 for the Div/P variable, 9 for the Debt/Equity variable, 10 for the growth in sales variable, and 11 for the volume variable.

Mathematically, if we use t'_i to denote the publication year of anomaly i, the values presented in these two figures are $RP1_{i,j}(t'_i, 2006)$ and $RP2_{i,j}(t'_i, 2006)$, where i ranges from 1 to 11 for different anomalies and j may equal 1,2,5, or 10 for different training lengths. Figure 2.5 is for all anomalies and Figure 2.6 is for the case when only cross-sectional return anomalies are considered. I report both the relative popularity measure that counts first place winners only (RP1) and the one that considers all ranking information (RP2). In Figure 2.5, as the training length increases from one year to ten years, the anomaly selections converge to the January effect—the best strategy in-sample; yet, the 10-year training length does not yield the highest real-time return, as we can see in Figure 2.3. In Figure 2.6, although the anomaly selection does not converge to the single best anomaly, the relative popularity values with the 10-year training length roughly reflect the performance ranks of these anomalies in-sample (e.g., the momentum strategy is the most "popular" and has the best in-sample performance).

As we can see from Figures 2.5 and 2.6, as well as Table 2.2, the January effect is the best performing anomaly in-sample and therefore has the highest relative popularity. Among cross-sectional anomalies, the momentum effect performs the best if we neglect the volume effect because of its late availability. This leads to the possibility that the whole result might be driven by these two anomalies. Table 2.6 presents the out-of-sample performance when excluding one or both of these two anomalies. As we remove one or two of the best anomalies from the trader's consideration, the out-of-sample performance decreases as expected. However, we still see a positive excess return and a positive Alpha, and they are significant in most of the other cases.

2.4.2 Are Old Anomalies Less Profitable?

Using the data of relative popularity for each anomaly in every year, I run fixed effects regressions to see if relative popularity measures, as well as the excess return, decrease as the "age" of an anomaly—the number of years after publication—increases. Specifically, I run the following three regressions on RP1, RP2, and the anomalous excess return to see if the number of years after publication has any effect on the values of these measures controlling for individual anomaly dummies. For a given training length j,

$$RP1_{i,j}^{t} = \gamma_{1,11} + \beta_1 Age_i^t + \Sigma_{m=1}^{10} \gamma_{1,l} I(m=i) + \varepsilon_{1,i,j}^{t} \quad (2.12)$$

$$RP2_{i,j}^{t} = \gamma_{2,11} + \beta_2 Age_i^t + \Sigma_{m=1}^{10} \gamma_{2,l} I(m=i) + \varepsilon_{2,i,j}^{t} \quad (2.13)$$

$$ER_{i,j}^{t} = \gamma_{3,11} + \beta_3 Age_i^t + \Sigma_{m=1}^{10} \gamma_{3,l} I(m=i) + \varepsilon_{3,i,j}^{t} \quad (2.14)$$

$ER_{i,j}^{t}$ represents the cumulative excess return over the market benchmark for

Table 2.6. Out-Of-Sample Performance without the January Effect and/or the Momentum Strategy

This table presents the out-of-sample performance of a trader who picks the best performing anomaly among published anomalies recursively over time. Panels A through D represent different training lengths of 1, 2, 5, and 10 years, respectively. I report three different anomaly groups represented by *Group Name*: *No Jan* is for all anomalies excluding the January effect; *No Momentum* is for all anomalies without the momentum strategy; *No Both* is for all anomalies excluding both the January effect and the momentum strategy. For each anomaly group, I report the average annual return (*Raw Return*) and the average annual benchmark return (*Return Bench*). I also report four performance measures of the real-time return over the benchmark index: the annual excess return (*Excess Return*), the one-factor *Alpha*, the difference in Sharpe ratios between the real-time trader and the market index (*Diff Sharpe*), and the difference of the certainty equivalent rate of returns between the real-time trader and the market index with a risk aversion level of 2 (*Diff CER*). The detailed calculations of *Diff Sharpe* and *Diff CER* are described in Section 2.2.2. The standard deviations are listed in brackets.

Group Name	Raw Return	Bench Return	Excess Return	Alpha	Diff Sharpe	Diff CER
	Panel A: 1-Year Training Period					
No Jan	18.19%	13.63%	4.56%*	4.31%	0.11	3.16%
	(19.48%)	(15.47%)	(12.64%)	(2.62%)	(0.29)	(4.6%)
No Momentum	18%	13.63%	4.37%**	5.85%**	0.23	4.25%
	(15.86%)	(15.47%)	(11.29%)	(2.26%)	(0.29)	(4.07%)
No Both	15.7%	13.63%	2.07%	3.69%	0.1	2.13%
	(15.27%)	(15.47%)	(11.22%)	(2.23%)	(0.31)	(4.18%)
	Panel B: 2-Year Training Period					
No Jan	23.79%	13.63%	10.16%***	9.25%**	0.25	7.15%*
	(23.25%)	(15.47%)	(16.9%)	(3.49%)	(0.29)	(4.96%)
No Momentum	23.17%	13.63%	9.54%***	10.69%***	0.36*	8.2%**
	(19.32%)	(15.47%)	(15.54%)	(3.19%)	(0.3)	(4.24%)
No Both	22.65%	13.63%	9.01%***	11.33%***	0.34*	7.8%**
	(18.99%)	(15.47%)	(17.47%)	(3.49%)	(0.32)	(4.31%)
	Panel C: 5-Year Training Period					
No Jan	24.88%	13.63%	11.24%***	11.32%***	0.33	8.86%**
	(21.86%)	(15.47%)	(16.98%)	(3.53%)	(0.29)	(4.67%)
No Momentum	24.25%	13.63%	10.61%***	12.04%***	0.42*	9.4%**
	(18.99%)	(15.47%)	(15.82%)	(3.23%)	(0.32)	(4.24%)
No Both	20.49%	13.63%	6.85%**	8.52%**	0.2	5.17%
	(20.19%)	(15.47%)	(17.44%)	(3.55%)	(0.29)	(4.42%)
	Panel D: 10-Year Training Period					
No Jan	19.47%	13.63%	5.84%**	5.79%*	0.13	3.99%
	(20.59%)	(15.47%)	(14.95%)	(3.1%)	(0.31)	(5.03%)
No Momentum	20.97%	13.63%	7.34%***	9.29%***	0.49*	7.61%**
	(14.55%)	(15.47%)	(10.83%)	(2.09%)	(0.35)	(4.04%)
No Both	18.69%	13.63%	5.06%*	7.79%**	0.24	4.78%
	(16.37%)	(15.47%)	(15.33%)	(2.97%)	(0.34)	(4.41%)

***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

anomaly i in year t for the past j years. Age_i^t is the number of years after publication for anomaly i as of year t. β_k is the coefficient on Age for the k^{th} regression. $\gamma_{k,i}$ is the loading on the i^{th} dummy variable for anomaly i in regression k. $\varepsilon_{k,i,j}^t$ is the White standard error which is robust to heteroscedasticity.

Table 2.7 reports the results for the above three regressions. As we can see, there is no significant attenuation over time for any of the three measures if the real-time trader looks back for one year, two years, or five years. There are some cases with significant decrease in performance over time if the training length is 10 years. However, these results are driven by pre-publication performances which are significantly better than those after publication. For example, if an anomaly is published in 1981, the real-time trader back testing it in 1982 using a 10-year training length would find a very good performance because for nine out of ten years in the training period the anomaly would not have been published yet. Overall, although anomalies experience decreased in-sample performance as documented in Schwert (2003) as well as in Table 2.2 of this paper, this reduction is not related to how long the anomaly has been published. An old anomaly is as likely to be selected by the real-time trader as a new one. Any performance reduction due to trader participation, if it occurs, happens immediately after the initial publication of the anomaly.

We could also infer from Table 2.7 that the performance of published asset-pricing anomalies is quite stable over time. Therefore, any dynamic trading strategies taking advantage of published anomalies, such as the one presented in this paper, are likely to perform similarly in the future. Overall, the empirical results in Table 7 indicate the future performance for the simulated real-time trader would remain similar to what we find using historical data.[17]

[17] Of course, here we neglect the impact of the circulation of this paper itself. As we have already seen, academic publication does (negatively) impact the performance of the anomaly. This implies that whether or not the results found here will hold in the future depends on the potential future publication of this paper.

Table 2.7. Relative Popularity, Excess Returns, and the Age of Anomaly

This table presents the results of the following fixed effects regressions on the relative popularity measures RP1 and RP2, as well as the annual excess return (ER), for the 11 basic anomalies.

$$RP1^t_{i,j} = \gamma_{1,11} + \beta_1 Age^t_i + \Sigma^{10}_{m=1}\gamma_{1,l}I(m=i) + \varepsilon^t_{1,i,j}$$
$$RP2^t_{i,j} = \gamma_{2,11} + \beta_2 Age^t_i + \Sigma^{10}_{m=1}\gamma_{2,l}I(m=i) + \varepsilon^t_{2,i,j}$$
$$ER^t_{i,j} = \gamma_{3,11} + \beta_3 Age^t_i + \Sigma^{10}_{m=1}\gamma_{3,l}I(m=i) + \varepsilon^t_{3,i,j}$$

Panels A through D represent different training lengths of 1, 2, 5, and 10 years, respectively. β_1, β_2, and β_3 are the coefficient for the RP1 measure, the RP2 measure, and the raw annual excess return, respectively. t statistics and R^2 based on the White standard error are reported for each coefficient with statistical significance marked.

Panel A: All Anomalies									
Training	β_1	t	R^2	β_2	t	R^2	β_3	t	R^2
One Year	0.0022	0.17	0.06	-0.0024	-0.76	0.17	0.0015	1.04	0.07
Two Years	0.0043	0.33	0.07	-0.0021	-0.66	0.25	0.003	0.87	0.16
Five Years	0.0044	0.35	0.28	-0.0022	-0.77	0.4	0.0018	0.25	0.32
Ten Years	0.0021	0.29	0.72	-0.0046**	-2.12	0.6	-0.0468**	-2.56	0.59
Panel B: No January									
Training	β_1	t	R^2	β_2	t	R^2	β_3	t	R^2
One Year	-0.0032	-0.25	0.06	-0.0033	-0.97	0.11	0.0017	1.02	0.06
Two Years	-0.0002	-0.02	0.11	-0.0029	-0.82	0.17	0.0035	0.92	0.13
Five Years	-0.0034	-0.26	0.2	-0.0034	-1	0.27	0.003	0.37	0.21
Ten Years	-0.0264***	-3.33	0.4	-0.0063**	-2.39	0.44	-0.0283**	-1.65	0.34
Panel C: No Momentum									
Training	β_1	t	R^2	β_2	t	R^2	β_3	t	R^2
One Year	0.0072	0.56	0.07	-0.0018	-0.55	0.18	0.0017	1.15	0.06
Two Years	0.0087	0.67	0.09	-0.0015	-0.45	0.26	0.0038	1.09	0.16
Five Years	0.0052	0.42	0.38	-0.0015	-0.51	0.39	0.0033	0.45	0.30
Ten Years	0.0014	0.21	0.81	-0.0041*	-1.85	0.59	-0.0521*	-2.78	0.62
Panel D: No January and Momentum									
Training	β_1	t	R^2	β_2	t	R^2	β_3	t	R^2
One Year	0.0056	0.47	0.05	-0.0026	-0.75	0.1	0.0019	1.13	0.05
Two Years	0.0113	0.87	0.06	-0.002	-0.55	0.16	0.0045	1.14	0.11
Five Years	0.0101	0.8	0.13	-0.0023	-0.68	0.23	0.0048	0.58	0.16
Ten Years	-0.0159*	-1.67	0.3	-0.0056**	-2.04	0.40	-0.0334**	-1.91	0.29

***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

2.5 Robustness Checks

The purpose of this research is to remove data-snooping bias incurred when researchers choose anomalies to test based on recent literature. By using out-of-sample results that are less subject to data-snooping bias, we can predict future trading profits with more confidence. However, this research itself still falls into the paradigm of empirical asset pricing, which relies on historical data to infer a model, and then uses the model to predict future returns. Whether or not the conclusions in this research will hold in the future still depends on unknown data that will be available in the future. In this section, I use different data lengths to check the robustness of the main results of this study. In order to do this, I repeat the whole set of analyses at points of time five years ago and ten years ago, and see if the main results still hold. I expect the results to remain qualitatively the same but with less statistical significance in these simulations because of the reduced sample size.

Another fact in the real world is the lag time between when a paper is forthcoming and when it finally appears in a top journal. A real-time trader could learn about an anomaly early when the paper is still in its forthcoming stage. To correct for this, I also repeat the whole set of analyses with anomalies available two years before each of the publication years. In this case, the simulation should generate better performance because the real-time trader could take advantage of anomalies at an earlier time. In addition, a test postponing the availability of anomalies two years after publication is also conducted here. This case corresponds to the possibility of slow information diffusion after academic publication. Anomalies are not always exploited immediately after they become known to the public.

In addition, I also test an alternative anomaly selection criterion, the Sharpe ratio, in this section. Instead of selecting the best anomaly based on past return, the real-time trader now picks the strategy with highest Sharpe ratio during the training period. To calculate Sharpe ratio, we need the monthly return data for each anomaly. Since the two calendar anomalies do not have well defined monthly return, I limit the test to those cross-sectional return anomalies by using the monthly portfolio returns to calculate training period Sharpe ratio.

Therefore, I repeat the real-time simulation for four additional scenarios: with data ending in 2001, with data ending in 1996, and with each publication year minus or plus two. Due to space constraints, I do not report all the results here. Instead, I list the following key conclusions about the real-time performance:

- The rank of training lengths

  So far we can see from Figure 2 that real-time profit is the highest when the trader chooses a 5-year training length, followed by 2-year, 10-year, and 1-year training lengths. I report the performance rank of training lengths for each robustness check scenario to see if the rank changes over time or not.

- The average performances using four performance measures

  In Table 3, I report the out-of-sample performances using four performance measures in 12 cases: four different training lengths each with three different anomaly groups. In this section, I recalculate the average performance of these 12 cases for each robustness check scenario.

- The empirical p-values

  Figure 6 reports the empirical p-values using four performance measures. I report these p-values for each robustness check scenario.

Table 2.8. Main Conclusions under Different Robustness Check Scenarios

This table compares major conclusions about the real-time profitability across five different scenarios. Columns 2006, 2001, and 1996 represent the situations when the available data ends in year 2006 (full data length), 2001, and 1996, respectively. "Minus Two" is the case when all the publication years of anomalies are subtracted by two to compensate for the time as a forthcoming paper. "Plus Two" is the case when publication years are added by two if the information diffusion process is slow. Rank of Training is the performance rank among different training lengths of one year, two years, five years, and ten years. Avg_Ex_Rt, Avg_Alpha, Avg_D_Sharpe, and Avg_D_CER are the average values for the four performance measures across the 12 cases presented in Tables 3. The four performance measures are the excess return measure, the one-factor Alpha measure, the Sharpe ratio difference measure, and the CER return difference measure. *p*_Ex_Rt, *p*_Alpha, *p*_D_Sharpe, and *p*_D_CER are empirical *p*-values for the excess return measure, the one-factor Alpha measure, the Sharpe ratio difference measure, and the CER return difference measure, respectively.

Scenario	2006	2001	1996	Minus Two	Plus Two
Rank of Training	5,2,10,1	5,2,10,1	5,2,10,1	5,2,10,1	5,2,10,1
Avg_Ex_Rt	8.10%	7.76%	5.87%	8.79%	7.30%
Avg_Alpha	8.65%	8.81%	6.50%	8.95%	8.40%
Avg_D_Sharpe	0.33	0.34	0.23	0.33	0.29
Avg_D_CER	6.89%	6.79%	5.10%	7.38%	6.20%
*p*_Ex_Rt	0.0298	0.0894	0.1553	0.0088	0.0166
*p*_Alpha	0.0078	0.0567	0.1192	0.0002	0.0078
*p*_D_Sharpe	0.1456	0.2490	0.4025	0.0596	0.1505
*p*_D_CER	0.0669	0.1666	0.2443	0.023	0.0508

Table 2.8 compares the main conclusions of this paper across five different scenarios: The current scenario with data ending in 2006; the 5-year-ago scenario with data ending in 2001; the 10-year-ago scenario with data ending in 1996; the forthcoming paper scenario with anomalies available two years before publication; and the slow information diffusion scenario with anomalies available two years after publication. As we can see, the results are similar across these scenarios. 5-year and 2-year training lengths consistently yield better out-of-sample performances than 10-year and 1-year training lengths. Just as we expected, knowing an anomaly early in its forthcoming stage improves the real-time trader's performance. In addition to better average performances for all of the four measures, the column "Minus Two" also shows lower p-values for all performance measures. As expected, the column "Plus Two" has results changed in the other direction. As we can see from Table 2.8, the significance of findings decreases as we switch the calendar back to 2001 and then to 1996 because of the reduced number of samples. However, the results remain qualitatively the same as the current scenario with the full length of data. In unreported tables where available data ends in 2001 and in 1996, as well as those where the publication years are set two years earlier(or later), I find similar results about the real-time profitability with a changing number of anomalies under consideration. Overall, simulation runs using truncated data suggest that the main findings in this paper would be the same had I conducted this research earlier. The alternative assumptions that an anomaly becomes available earlier or later than its publication year do not change our main conclusions either.

Finally, in Table 2.9 I report the out-of-sample performance when the real-time trader uses the Sharpe ratio as the alternative anomaly selection criterion. The results are for the situation when the trader only considers the nine cross-sectional return anomalies because they have well-defined monthly portfolio returns. Compared with the performance using past returns as the strategy selection criterion in Table 2.3, the case with one-year training length has a better result. The five-year and ten-year cases remain roughly the same, but the two-year training length yields a worse result when the Sharpe ratio criterion is used. Overall, we can still see positive excess performance over the market benchmark in all cases using all four measures. The result is robust to the alternative Sharpe ratio strategy selection criterion.

Table 2.9. Out-of-Sample Performance Using Sharpe Ratio as Anomaly Selection Criterion

This table presents the out-of-sample performance when the real-time trader uses Sharpe ratio as the alternative anomaly selection criterion. The results are for the nine cross-sectional return anomalies only. I report results with different training lengths in each row. For each training length, I list the average annual return (*Raw Return*) and the average annual benchmark return (*Return Bench*). I also report four performance measures of the real-time return over the benchmark index: the annual excess return (*Excess Return*), the one-factor *Alpha*, the difference in Sharpe ratios between the real-time trader and the market index (*Diff Sharpe*), and the difference of the certainty equivalent rate of returns between the real-time trader and the market index with a risk aversion level of 2 (*Diff CER*). The detailed calculations of *Diff Sharpe* and *Diff CER* are described in Section 2.2.2. The standard deviations are listed in brackets.

Training Length	Raw Return	Bench Return	Excess Return	Alpha	Diff Sharpe	Diff CER
One Year	20.26%	13.63%	6.63%**	5.53%*	0.13	4.22%
	(21.91%)	(15.47%)	(14.6%)	(3%)	(0.28)	(4.54%)
Two Years	16.83%	13.63%	3.2%	2.61%	0	1.11%
	(21.18%)	(15.47%)	(14.31%)	(2.96%)	(0.28)	(4.36%)
Five Years	17.57%	13.63%	3.94%**	6.36%***	0.33	4.5%
	(13.53%)	(15.47%)	(10.49%)	(1.92%)	(0.34)	(3.94%)
Ten Years	14.91%	13.63%	1.28%	4.31%*	0.12	1.77%
	(13.78%)	(15.47%)	(12.65%)	(2.29%)	(0.33)	(3.96%)

2.6 Conclusion

This paper examines the question whether or not a trader can beat the market in real time by picking the best published asset-pricing anomaly. To reduce the potential data-snooping bias in the out-of-sample test of asset-pricing anomalies, I rely on the academic publication process to reveal candidate anomalies for the consideration of a real-time trader, rather than determine the anomalies to be tested based on recent literature reviews. My conclusion generally shows that by taking advantage of published anomalies, a trader can beat the market. This result is stronger when the trader looks at past performance between two years and five years, and is robust to several exogenously determined variables such as training length and the group of anomalies under consideration. However, caution needs to be taken in interpreting this result as a prediction into the future. Falling into the same empirical research paradigm, this research relies on past performance data to draw inferences about future profitability.

As in many other anomaly papers, one final question is: Why can the trader beat the market in real time? The methodology applied in this paper rules out the following two possible reasons: 1) The good performance is the result of data snooping; and 2) traders would not know these anomalies in real time. However, before we could conclude the market is inefficient, the impact of the limits of arbitrage and the trader's learning behavior needs to be carefully considered. Specifically, the trader may have only limited ability to take advantage of arbitrage opportunities if 1) his money under management is positively correlated with his short-term performance, as is illustrated in Shleifer and Vishny (1997), *or*, 2) the trader is skeptical about the ability to beat the market so he Bayesianly allocates funds between the active strategy and the market portfolio as in, for example, Brennan and Xia (2001).

An interesting example of how short term performance fluctuation may affect the profitability can be found in this study with the case using a one-year training period and all published anomalies. In this case, the five best years in our simulation ranked from the best are 1999, 1977, 2001, 2002, and 1991, during which the real-time trader beats the market by large margins, while the five worst years, starting from the worst, are 1989, 1988, 2000, 1980, and 1998. If the real-time trader or his mutual fund clients learn from recent performance, the trader would miss the best year of 1999 because of the bad performance in 1998. If he corrects this mistake after observing the good performance in 1999, he would run into another rough year of 2000. Based on the past year performance, the real-time trader would not pick the right side until 2002 when both the previous and the current performances are good. However, would the real-time trader (or his investors) change this learning behavior at the end of 2001 after being fooled by past performances for three consecutive years? Of course, because this example bears no statistical significance, it just illustrates the difficulty and the complexity of real world investment decisions.

Chapter 3

Managerial Decisions and Mutual Fund Performance

3.1 Introduction

Empirical evidence shows that the difference between firms' earnings reports and the earnings forecasts demonstrates a negative-skewed distribution with a dent to the left (Burgstahler and Dichev, 1997). This means that earnings reports tend to beat forecasts a little bit while those who cannot meet the forecast end up far below their targets (i.e., earnings forecasts in the accounting context). The accounting literature attributes this phenomenon to the manager's manipulation of the accounting earnings, that is, earnings smoothing or a "big bath" (Kirchenheiter and Melumad, 2002, KM here after). This manipulation makes the true economic earnings of a firm in general different from the reported earnings. In the KM model, the manager's purpose is to surprise the investors in order to get a higher payoff. Because the payoff function in terms of the degree of surprise is concave, managers have incentive to report slightly better-than-expected reports from time to time by using earnings smoothing. However, the punishment on negative surprises is convex, motivating the manager to report very low accounting earnings if she sees a bad economic earnings that could not be smoothed to a positive increment. Another theoretical approach on the skewed earnings distribution is by Guttman, Kadan and Kandel (2003) using the rational expectations framework. In their model, the manager's payoff is related to the firm's stock price, which in turn is determined by investors' rational expectations on the firm's true value. They show the existence of a pooling equilibrium in which the manager adds some "home-made" noise to the earnings report.

While it is reasonable to propose earnings manipulation explanations because investors can only infer the true value of the firm through accounting reports, in many other areas where the true value of a risky activity is directly observable, the results still demonstrate such a negative-skewed distribution. For example, test scores for students usually show a "double-hump" distribution with most students doing ok or above average while a small group of them fall far below the standard norm. This can be explained as follow: Students feeling good about the exam tend to be conservative in the final preparation. For examples, they do not stay up late in the night before the exam and during the exam they do not spend too much time on one question because they know they are going to get most of the questions. On the other hand, those students slightly below the standard norm take some risky actions in the hope to reach their goals. For examples, staying up late before the exam or shooting at some difficult questions even there is not enough time during the exam. And those far below the standard are giving up so they do not take risky actions either. In this example, there is no manipulation in the final result—the test score. Then, the research question in this paper is: Can the negative-skewed distribution arise without earnings management?

The answer to this question partly comes from the mutual fund literature. Obviously, fund managers cannot "manipulate" mutual fund returns. However, facing a similar pay-performance relation as firm managers do, mutual fund managers have other ways to change the distribution of terminal fund values.

For example, Chevalier and Ellison (1997) find that fund managers increase the riskiness of their equity portfolios from September to December. Brown, Harlow and Starks (1996) find the similar results in a tournament setting assuming a manager's payoff is related to the rank of fund performance relative to other funds. Theoretically, Grinblatt and Titman (1989) first model the managerial behavior facing performance-based contracts. They show risk neutral managers take excessive risks when there is an option-like payoff scheme. Recently, Carpenter (2000) demonstrate that a risk averse manager will also deviate from the optimal risk level, sometimes lower to "lock in" the profit and sometimes higher to take advantage of the option-like payoff. Therefore, a possible empirical question is if we can find mutual fund performance to have the similar distribution as that of accounting earnings.

This paper proposes a simple model where the risk averse manager can choose whether or not to take a fixed level of uncertainty in the form of a risky project in the final stage of a managerial term. In addition to a linear payoff related to firm's terminal value, the manager has a reward for beating a break even point, or a benchmark. By beating the benchmark, the manager acquires additional personal satisfaction represented by a dollar amount bonus. When the bonus is sufficiently large to compensate for the risk of taking the project, the manager chooses to take the risk if she is slightly below the target, but not to take it if she is above the target or far below it. The manager's decision based on the random realization of the true economic earning makes the terminal value of the company to be likely either above the benchmark, or far below it. This risky decision also makes the distribution of firm's terminal value a negative-skewed distribution *without* earnings management.

In this model, there are two important assumptions. First, unlike the existing literature on managerial risk behavior, this paper assumes a linear (symmetric) payoff for managers with respect to firm's terminal value. This is partly because from the regulatory side, mutual fund managers are prohibited from the asymmetric option-like pay-performance structure.[1] Also, as documented in Chevalier and Ellison (1997) and Gruber (1996), fund assets under management are positively related to performance. Therefore, it is reasonable to set a linear compensation scheme for fund managers when their pay is based on assets under management. Second, while a linear payoff alone will totally eliminate the possibility for the risk averse manager to take any extra risks, I assume the manager has a behavioral bias towards reaching a benchmark, which generates the convexity in utility function necessary for a possible risk seeking behavior. This tendency to reach a psychological break even point, referred as mental accounting, is well documented in behavior finance literature such as Thaler (1985) and Thaler and Johnson (1990). To account for the psychological bias of meeting or beating a target, I assume a dollar amount bonus for managers to achieve the goal, just as in Degeorge, Patel and Zeckhauser (1999).

I test this model using the performances of growth mutual funds from 1994 to 2004. Consistent with Chevalier and Ellison (1997), I find that funds have a higher variance on excess returns for the 4th quarter of the year, indicating that fund managers on average take more risks near the year end to meet or beat a performance benchmark. However, this result needs to be further verified because the 4th quarter high variance is mainly from years 1999 and 2000. I also find persistence on the one year lag performance as documented in Carhart (1997). To test the hypothesis that slightly below average funds take more risk, I focus on the funds with mediocre performance rather than those with extreme performance.[2] I find that funds in the 3rd and 4th worst deciles on average outperform funds in the 5th and have higher return variances in the next year. This result is robust when

[1]But this restriction does not apply to hedge funds. See Das and Sundaram (2002) for an analysis of possible fee structures.

[2]I also find that funds with extremely low and extremely high past year performance have higher excess return variance. This could be explained by Elton, Gruber and Blake (2003) who find that good performance is at the cost of high risk exposure. The high volatility for low performance funds may be best explained by the theory of option-like payoffs modeled by both Grinblatt and Titman (1989) and Carpenter (2000).

funds are partitioned into duo-deciles. I also find the excess returns for growth funds are more positively skewed than those for index funds. This evidence does not support my model in the mutual fund context, but could be explained in a multi-period model in which good managers keep on trading profitably while losers cease trading after small losses as modeled in Mahani and Bernhardt (2007).

The rest of the paper is arranged as follow: Section 2 reviews the related literature. Section 3 is the model setup. Section 4 solves the individual risk choice problem with discussion on the comparative statics. Section 5 illustrates the distribution of firms' terminal value. Section 6 shows some empirical evidence on mutual fund performance. Section 7 is the conclusion.

3.2 Related Literature

My model is similar to that of Carpenter (2000), which studies the risk preference of a risk averse manager with stock option compensation. However, my model is different in two aspects regarding the assumptions on manager's payoff. First, with respect to the terminal value of the firm, my model assumes the manager has a linear payoff rather than a convex option-like payoff. This is a stronger assumption because under a linear payoff scheme it is more difficult to generate the risk-seeking behavior of managers than under an option-like pay-performance schedule. One reason could be on the mutual fund regulatory side, option like pay-performance compensation is prohibited. Rather, the manager has to face a symmetric "fulcrum" style compensation. That is, while rewarded for a better-than-average performance, the manager will also be punished in the case of unsatisfactory performance. This assumption makes the risk averse manager less willing to take risks because there is no protection from the downside. Second, my model assumes if the fund performance meets a performance benchmark, there exists a bonus to the manager in the form of a step function of performance. This accounts for the psychological satisfaction of reaching a break even point. In the real world, maintaining a satisfactory level of performance will not only make the manager happy but also secure the fund manager's job.

I also get some new results comparing with Carpenter (2000). One finding is that if the bonus cannot compensate for the utility loss of risk aversion for the manager just slightly below the break even point, no manager will take additional risks at all. This reflects the intuition that if the risk of a project cannot be flexibly adjusted, risk averse manager may just pass it even if it has positive NPV for the manager or investors.

This paper is also related to the work of Grinblatt and Titman (1989), which also has a bonus fee to generate the covexity of manager's payoff. They assume the manager is risk neutral. Instead, this paper has a stronger assumption that the manager is risk averse. Therefore, the assumptions needed for the managerial risk-seeking behavior is minimized.

While most of the papers on managerial risky decision show managers may take too much risk, Admati and Pfleiderer (1997) argue that privately informed managers could end up bearing too little risk by indexing the market, because they want to secure the incentive fee benchmarked on an index. In my model, I do not assume managers have inside information simply because empirical evidence does not find fund managers have superior stock picking ability substracting the management fees.[3] Moreover, the result of Admati and Pfleiderer (1997) can be incorporated as a special case in my model where the risk of deviating from a linear payoff is too huge relative to a profitable project.

Empirically, this paper finds mutual fund managers take additional risks especially if their performance is slightly below a benchmark. This finding adds up to the finding of Chevelier and Ellison (1997) that mutual fund managers take more risks later in a year. There is also opposite evidence from professional

[3]For evidence on this point, see for example, Elton, Gruber, Das, and Hlavka, (1993)

traders. Coval and Shumway (2005) show that Chicago Board of Trade (CBOT) proprietary traders take more risky positions during afternoon trading if they win in the morning, and are less risky if they lose in the morning. As their paper mentions, this behavior is largely attributed to traders behavioral biases such as the house-money effect and the loss aversion. In addition, relative to the short horizon, daily repeating decision game of a day trader, the manager in my model makes longer horizon and less frequent decisions that are less subject to behavioral biases.

3.3 Model

The model in this paper is a two-period, one-agent(representative) principal-agent problem with a linear payoff of the terminal value of the firm plus a fixed bonus utility if the firm's time-2 value is above a target. A single optional risky project can be chosen by the manager at time-1 based on her utility maximization problem. Below are the details of the model:

- The Firm

  The firm in this model has a normally distributed initial value $\widetilde{m}$ realized at time 1,[4] with a mean of v and variance S_m, that is, $\widetilde{m} \sim N(v, S_m)$. The random initial value is to capture the diversification of firms' status by using only one representative firm. For modeling convenience, I set the mean of firm initial value v to be exactly the expectation, or benchmark for the manager. In other words, the benchmark in this model is the social average performance.

- The Risky Project

  There is a single risky project that can be taken at time 1 which eventually

[4] In this paper, we use tilde "$\widetilde{\ }$" to represent random variables.

pays off $\widetilde{x} \sim N(0, S)$ to the firm at time 2. In order to simplify, I set the expected return for the project to be zero. However, the main findings in this paper will not be affected if adding a positive mean to the return. Finally, for the manager at time 1, the terminal value of the firm will be: $\widetilde{y} = m + \widetilde{x}$ if the manager takes the risky project and $y = m$ if she does not.

- The Agent

  There is one representative manager with constant absolute risk aversion (CARA) utility function:

$$U(w) = 1 - \exp(-\alpha w) \tag{3.1}$$

  where α is manager's coefficient of absolute risk-aversion and w is her payoff at time 2. At time 1, the manager observes the realization of the firm's initial value m and decides whether to take the risky project or not. Let $e = 1$ denote that she takes the project and $e = 0$ if not. Her time-2 payoff is contracted linearly on the realization of firm's time-2 value: y. For simplicity, I set this compensation to be $w = ky$. Furthermore, given that the firm value can be measured in different units in the real world, without loss of generality, we can assume $k = 1$. Plus, she will get a bonus B in dollar amount and $U(B) = 1 - \exp(-\alpha B)$ in utility if the terminal value is above the expectation, that is, if $y \geq v$.

3.4 The Individual Risk Choice

In this section, I solve for the manager's utility maximization problem at time 1 as an optimal decision rule on whether to take the risky project or not. Then, I analyze the comparative statics on model parameters.

The utility maximization problem is trivial for the manager if the realized time-

1 firm value is already greater or equal to the expectation. In this case taking the risky project will not only decrease her utility due to risk aversion, but also incur greater uncertainty of getting the bonus utility that otherwise would be given for sure, that is, $E[U|e = 0] = U(m) + U(B) > U(m - (1/2)\alpha S) + Pr(m + x \geq v)U(B) = E[U|e = 1]$.[5] Here $Pr(.)$ means the probability of something to be true. Therefore, for realizations with $m \geq v$, the manager will not take the risky project and the terminal value of the firm will stay the same as the initial value: $y = m$.

If the realization of firm's initial value is: $m < v$, the expected utility for the manager is:

$$E[U] = \begin{cases} U(m) & \text{if not taking it, } e = 0 \\ U(m - (1/2)\alpha S) + Pr(m + x \geq v)U(B) & \text{if taking it, } e = 1 \end{cases} \tag{3.2}$$

Based on the expected utility, the time-1 manager's optimization problem can be summarized with the following proposition:

Proposition 1. *When* $[exp(-\alpha v)]\,[exp((1/2)\alpha^2 S) - 1] < (U(B)/2)$, *there exists a unique threshold* m^* *such that:*

1. *For all realizations of* m*:* $m^* < m < v$, *the manager takes the risky project. Otherwise, she does not take it.*

2. m^* *satisfies:*

$$\frac{1}{U(B)}\left(e^{-\alpha m^*}\right)\left(e^{(1/2)\alpha^2 S} - 1\right) = \Phi\left(\frac{m^* - v}{\sqrt{S}}\right) \tag{3.3}$$

where $\Phi(z) = \int_{-\infty}^{z}(1/\sqrt{2\pi})exp(-(u^2/2))du$ *is the normal cumulative density function and* $U(B) = 1 - \exp(-\alpha B)$.

[5] $U(m - \frac{1}{2}\alpha S)$ is the risk neutral equivalent utility given CARA utility under normal distribution. A classical reference is Grossman and Stiglitz (1980).

When $[exp(-\alpha v)]\,[exp((1/2)\alpha^2 S) - 1)] \geq (U(B)/2)$, *the manager does not take the risky project for any realization of* m.

Proof: See Appendix.

The condition of the existence of a threshold implies the following: For the special case where the initial firm value is almost but still slightly below the expectation v denoted by v_- (i.e., $m = v_-$),[6] the expected utility of getting the bonus by taking the project (with chance $\approx \frac{1}{2}$) must be greater than the utility loss caused by the risk of the project on a linear payoff scheme. When the two values are equal, we have: $U(v_-) - U(v_- - (1/2)\alpha S) = (exp(-\alpha v))\,(exp((1/2)\alpha^2 S) - 1) = (1/2)U(B)$ which is the existence condition for the threshold. Intuitively, this condition says the bonus, the risks of the project, and the risk averse coefficient should at least make those slightly-below-average managers willing to take the risk, otherwise nobody would take it at all. When the bonus (B) is too low, the risk of the project (S) is too high, or the manager is very risk averse (α is high), there could be no region for the realizations of m in which the manager takes the risky project. When the existence condition is not satisfied, our model reduces to a case similar to the result of Admati and Pfleiderer (1997) (AP1997 here after). Even the manager has a chance to get the bonus (she is informed as in the AP1997 model), she passes it and sticks to the current status quo (indexing as in AP1997) because of risk aversion.

Another interesting implication of the existence condition is that when the decision maker is risk-neutral ($\alpha = 0$), there is always a region (m^*, v) where she wants to take the risky project. We can fit this scenario to the case of outside equity holders, who are well diversified and therefore risk neutral. There might also be an incentive either psychologically or monetarily to the outside investors to break even at some point. Then, according to this case they would want the

[6] As a formal definition for v_-, it is a number satisfiying $0 < v - v_- < \varepsilon$, where ε is a sufficiently small positive number.

firm to take a risky positive NPV project if the firm value is slightly below the bar. In this situation the risk averse manager will pass a small positive NPV project, offering an alternative explanation to the under-investment problem of firms.[7]

When the existence condition is satisfied, the threshold m^* is determined by the first order condition when the utility loss caused by risk aversion is exactly offset by the expectation of getting the bonus. In the region $m \in (m^*, v)$, the manager has a very good chance of getting the bonus if she takes the risky project. In the region $m < m^*$, the chance of getting the bonus is too small comparing with the utility loss caused by payoff uncertainty. Therefore the manager will not take it. Figure 1 illustrates the existence and the uniqueness of the solution by plotting the left hand side (LHS) and the right hand side (RHS) of equation 3.3 in terms of m.

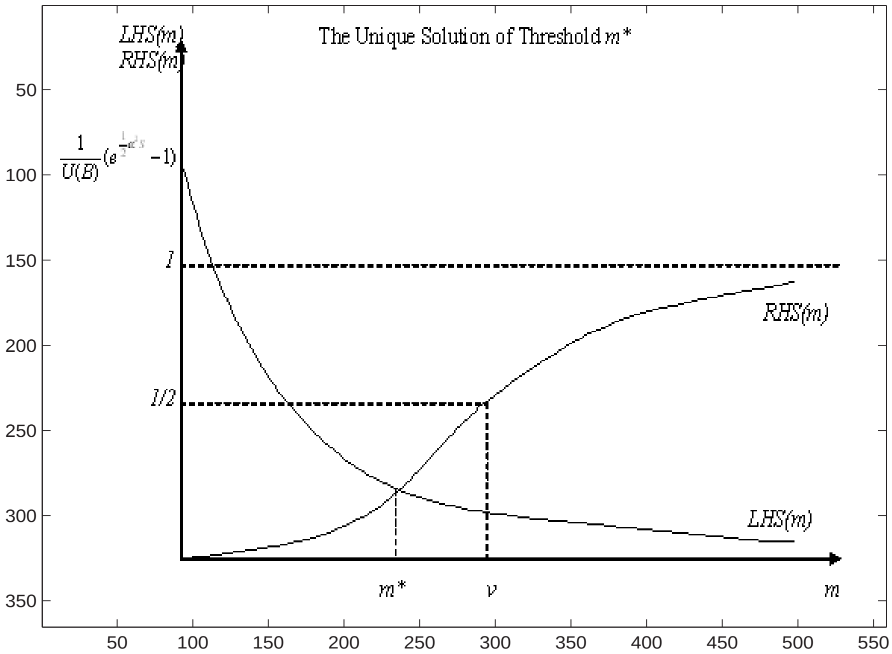

Figure 3.1. The Existence and the Uniqueness of the FOC Solution

The threshold m^* is affected by the exogenous parameters in several aspects,

[7]Other explanations in the literature include, for example, the bondholder-shareholder conflict documented by Parrino and Weisbach (1999)

as summarized in the next corollary.

Corollary 1. *When the existence condition is satisfied, m^* decreases as bonus B increases. The relation between m^* and S, and the relation between m^* and α are undecided.*

Proof: See Appendix.

When the satisfaction of meeting the expectation (B) is high, the manager is willing to take the risk under a large range of realizations of the firm initial value m. Therefore, the threshold m^* becomes lower.

However, the effects of the risk of the project (S) and the risk averse level (α) are mixed. When S increases, on the one hand the probability of getting the bonus increases, but the increased volatility also reduces the utility of a risk averse manager. The latter effect is caused by the linear payoff assumption. If the payoff is option-like, increased volatility will not affect manager's utility, therefore a riskier project is always more welcome.

For α, its relation to m^* is more complicated because α is not only related to all the utility functions, it also determines the gap between the initial utility and the risk neutral equivalent utility after taking the risky project. There are two competing effects determining the relation between risk averse level α and the reluctance to take risks m^*. First, the risk neutral equivalent payoffs of taking or not taking the project are: $m^* - (1/2)\alpha S$ and m^*. When α increases, there is a larger utility loss from taking the risky project, making the manager less willing to take it. This effect will push the threshold m^* towards the threshold v. Second, any deterministic payoff will seem more attractive to risk averse managers when α increases, because of the monotonicity of the CARA utility function on α. In this case, the bonus of reaching a break even point is more attractive to the manager, making her more willing to take the risky project and push the threshold m^* away from the threshold v. However, if we assume the bonus for beating a target (B) is far smaller relative to the performance based compensation for a manager, the former effect dominates. This leads to a higher threshold m^* that is necessary to trigger the risk taking activity for a more risk averse manager.

3.5 Terminal Value Distribution

This section discusses the realized distribution of firm's terminal value.

Proposition 2. *When a threshold m^* exists, the firms terminal value y has a distribution with the following density function:*

$$h(y) = I((y \leq m^*)or(y \geq v))\phi\left(\frac{y-v}{\sqrt{S_m}}\right) + \int_{m^*}^{v} \phi\left(\frac{y-x-m}{\sqrt{S_m}}\right)\phi\left(\frac{x}{\sqrt{S}}\right)dx \quad (3.4)$$

where $\phi(z)$ is the standard normal density function and $I(\cdot)$ is an indicator function.

Proof: See Appendix.

The probability density function in proposition 2 indicates a hump to the right and a fat tail to the left for the terminal value of the firm. This is caused by the probability "spill over" from the region of (m^*, v). The first term in equation (3.4) is the probability from the original normal distribution of the firm initial value m. The second term is the probability spilled over from the region where managers take the risky project. With the choice of a single risky project for the manager, the firms terminal value is likely to either meet the expectation target v or fall far below it. This is a result consistent with the earnings smoothing and big bath theory but without earnings management. Figure 3.2 illustrates a numerical example that shows the terminal distribution, given a possible solution of m^*.

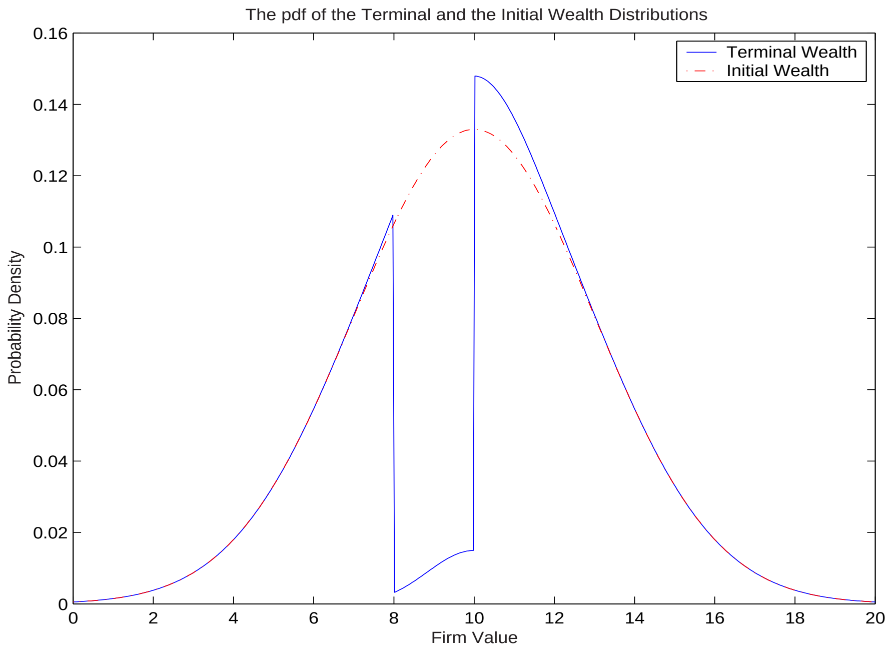

Figure 3.2. An numerical example of the Terminal Wealth Distribution, where $m = v = 10$, $m^* = 8$, $S_m = 3$ and $S = 1$.

3.6 Empirical Test on Mutual Fund Performance

The empirical part of this paper focuses on three aspects of mutual fund performance.

- The first part is the excess return distribution of actively managed funds and index funds. It answers the question: Do actively managed funds show additional negative skewness and leptokurtosis as suggested by the model in this paper? The answer is no.

- The second part is risk changes within a year. I address the question if there is more return variations towards the end of the year. The answer is yes but possibly caused by outliers.

- The last part is about the return distribution for fund deciles sorted by last year's performance (similar to Carhart, 1997). I want to see if those slightly below average managers take more risks as proxied by a larger return variation in the next year. The answer is yes.

Next I present these results following an introduction to the data.

3.6.1 Data

For the empirical research on the risk changes of mutual funds, I use the survivorship-bias free US mutual fund data from the Center for Research in Security Prices (CRSP). This data set has no survivorship bias because it includes all funds that were alive at a given year, despite whether or not they still exist today. I pick all the available funds from January 1993 to December 2004 which were alive at that time. Since I need to calculate the one year lag changes on variables such as the change of annual excess return, the 1993 data is only used to calculate lag variables for 1994. If a fund starts later than 1993, I only use its data from the second full operation year because the first full year is used for calculating lag variables. After cleaning missing data, there are 36657 fund-years out of a total of 51121 fund-years for growth funds. The data set includes 1054 funds in 1994 and the number of funds keeps increasing in the following years. I look at growth funds particularly because they are actively managed mutual funds that best fit into my model. Using the same procedure, I also get 10856 fund-years for index funds from 1994 to 2004 as a reference group in the analysis of return distributions.

The growth funds used in this research are marked as growth fund in CRSP, with ICDI's fund objective code to be either "growth" or "growth and income". The index funds are extracted from CRSP using a index funds list with 1622 tickers from investors fast track company (http://www.fasttrack.net). I calculate the gross quarterly and annual returns by compounding the monthly returns accordingly. Then I subtract the gross market returns proxied by the S&P500 returns to get the fund excess returns. I also sort funds within each year into 10 performance deciles. Slightly different from other papers such as Carhart (1997), I rank the fund performance in ascending order. That is, decile 1 funds perform the worst and decile 10 funds do the best. This ranking will make it easier for us to study the distribution of returns.

Table 3.1: Mutual Fund Data Summary Statistics (1994–2004)

This table is for the mutual fund data from 1994 to 2004. Turnover Ratio is the annual aggregate purchase or aggregate sell, whichever is smaller, divided by the net asset value of the fund. Expense ratio is portion of investment that goes into fund's expense. Percentage of Common Stocks is the money invested in common stocks as of the date. Gross return is the annual compounded return without adjusting for market return. Percentage still alive is the percentage of funds that still operate by the end of 2004. Panel A, B and C are for Growth Funds. Panel A is the summary statistics grouped by year. Panel B groups the data into current performance deciles in each year, with the lowest decile having the lowest excess return. Panel C groups all funds based on last year's return. We acquire the excess return data from year 1993 to get the last year deciles for 1994. Panel D is the summary statistics for index funds grouped by year.

Panel A: Summary Statistics by Year for Growth Funds						
Year	# of Obs.	Avg Turn Over	Avg Expense Ratio	Avg Pct of Common Stocks	Avg Gross Return	Pct Still Alive
1994	1054	76.29%	1.34%	89.07%	-1.93%	68.50%
1995	1393	81.70%	1.38%	89.65%	28.60%	69.99%
1996	1768	85.78%	1.44%	90.91%	17.98%	69.74%
1997	2150	83.51%	1.45%	91.69%	21.14%	70.05%
1998	2712	86.95%	1.45%	92.18%	14.20%	73.01%
1999	3272	90.13%	1.47%	91.24%	30.35%	75.86%
2000	3866	105.46%	1.48%	90.89%	-4.09%	78.38%
2001	4236	110.68%	1.48%	90.56%	-11.85%	82.41%
2002	4923	108.43%	1.56%	90.79%	-22.29%	85.90%
2003	5257	100.67%	1.58%	91.37%	32.88%	92.62%
2004	6026	92.71%	1.54%	91.54%	12.41%	99.17%
Total:	36657	97.01%	1.50%	91.09%	8.86%	83.18%
Panel B: Summary Statistics by Current Performance Deciles for Growth Funds (ascending)						
Current Decile	# of Obs.	Avg Turn Over	Avg Expense Ratio	Avg Pct of Common Stocks	Avg Gross Return	Pct Still Alive
1(low)	3621	130.73%	1.83%	90.65%	10.27%	73.93%
2	3700	107.14%	1.57%	92.88%	-2.89%	78.05%
3	3639	94.83%	1.56%	91.96%	0.74%	79.42%
4	3679	84.11%	1.42%	91.01%	4.10%	81.03%
5	3633	85.23%	1.40%	90.50%	6.57%	83.24%
6	3648	81.48%	1.38%	90.14%	8.83%	84.13%
7	3717	91.59%	1.46%	89.46%	11.72%	85.53%
8	3716	89.22%	1.43%	91.36%	15.93%	86.63%
9	3703	91.29%	1.41%	91.58%	20.92%	89.04%
10(high)	3601	115.08%	1.54%	91.35%	32.93%	90.78%
Total:	36657	97.01%	1.50%	91.09%	8.86%	83.18%

Table 3.1 continued:

Panel C: Summary Statistics by Past Performance Deciles for Growth Funds (ascending)						
Past Decile	# of Obs.	Avg Turn Over	Avg Expense Ratio	Avg Pct of Common Stocks	Avg Gross Return	Pct Still Alive
1(low)	3423	133.34%	1.85%	90.55%	4.19%	75.52%
2	3550	109.14%	1.59%	92.46%	6.39%	78.87%
3	3600	103.10%	1.56%	91.03%	8.00%	80.86%
4	3626	90.88%	1.45%	91.28%	7.69%	80.61%
5	3653	82.82%	1.41%	90.74%	7.20%	82.78%
6	3695	86.69%	1.37%	90.38%	8.26%	83.95%
7	3728	86.05%	1.45%	89.96%	9.87%	84.87%
8	3774	84.68%	1.42%	91.30%	10.92%	85.27%
9	3776	86.72%	1.40%	92.11%	11.65%	88.48%
10(high)	3832	109.88%	1.52%	91.07%	13.62%	89.38%
Total:	36657	97.01%	1.50%	91.09%	8.86%	83.18%

Panel D: Summary Statistics by Year for Index Funds						
Year	# of Obs.	Avg Turn Over	Avg Expense Ratio	Avg Pct of Common Stocks	Avg Gross Return	Pct Still Alive
1994	433	67.26%	1.09%	81.21%	-0.27%	93.87%
1995	504	76.11%	1.14%	81.45%	30.29%	94.47%
1996	574	79.19%	1.16%	83.57%	21.49%	94.81%
1997	756	77.42%	1.13%	83.56%	26.15%	95.75%
1998	921	78.59%	1.14%	85.96%	14.26%	96.47%
1999	1064	84.14%	1.16%	85.65%	22.36%	96.97%
2000	1151	103.26%	1.19%	86.45%	2.60%	97.75%
2001	1259	101.69%	1.15%	84.41%	-4.70%	98.49%
2002	1350	92.65%	1.18%	87.34%	-19.19%	98.91%
2003	1402	98.78%	1.19%	87.15%	33.96%	99.42%
2004	1442	87.05%	1.19%	86.80%	13.76%	99.65%
Total:	10856	89.06%	1.16%	85.38%	11.37%	97.59%

Table 3.1 summarizes the data. Panel A groups samples of growth funds by calendar years. We can see a steady growth in the number of funds over time. The ratio of funds still alive is also increasing as expected. These growth funds have relatively stable common stock holdings, but the turnover ratio is high around the burst of dot com bubble from year 2000 to 2003. The expense ratio also grows over time. Panel B summarizes the data categorized by current performance deciles. There are patterns on the expense ratio and turnover rate that funds at the two extremes tend to have higher expense and turnovers. This finding is exactly the

same as in Carhart (1997) who uses data before 1994. Panel C repeats the summary statistics grouped by last year's rank. We can see persistence in performance as well. There is an apparent trend of increase in the number of funds for each decile, indicating that fund survival is positively correlated to its performance. Panel D repeats the same statistics grouped by year as in Panel A for index funds. We see lower expense ratios, lower turnover rates, better performances, and a higher survival rate for index funds. However, due to the limitation of data access, the index funds list from fasttrack.net may suffer from selection bias that only the famous funds will be listed. Nonetheless, the index funds used in this research can provide a good reference on the higher orders of return distribution.

3.6.2 Return Distributions of Actively and Passively Managed Funds

The model in this paper suggests that fund managers slightly below a commonly accepted benchmark will take additional risks to meet or beat a target, making the terminal distribution of fund performance either slightly above or far below the benchmark. This is a distribution with negative skewness and leptokurtosis (fat tail). Therefore, the first test focuses on the skewness and kurtosis of the excess returns of actively manage mutual funds.

However, it should be noticed that the returns of a company may also hold characteristics on skewness and kurtosis. For some industries, doing business is usually ok but may occasionally suffer from terrible failures, such as pharmaceutical companies. In this case, funds holding these firms will show negative skewness in terminal distribution even without active management. On the other hand, in some risky industries such as high-tech or financial industry, the majority do not do very well, but some lucky few become super stars. This will make the distribution positively skewed. In both cases, the risk in real business should increase the kurtosis of the return distribution.

To separate the higher order distribution characteristics between firm level and fund level, I use passively managed index funds as a reference. I measure the skewness and kurtosis of those actively managed funds samples and compare them with values for index funds to see the effect of management behavior.

Table 3.2 shows the difference between index funds and growth funds in three aspects: return variation, skewness, and kurtosis for different time intervals. I use bootstrapping to see the significance of the difference. Just for verification purpose, the first three columns in Table 3.2 show that index funds have less return variation than growth funds by having statistically significant smaller standard deviations across all time frequencies. For skewness, the monthly and quarterly excess return show mixed evidence but in terms of annual excess return, both index funds and growth funds have positive skewness and the growth funds have higher positive skewness. There are leptokurtosis for both index funds and growth funds. Although index funds have thinner tails in many months and two quarters than that for growth funds, on the annual case index funds show even higher leptokurtosis than growth funds.

The empirical results in Table 3.2 do not match the predictions in the theoretical model of this paper. Instead, the evidence supports the recent theoretical work by Mahani and Bernhardt (2007), which models the financial speculation as a multi-period learning process. In their model, losers cease or significantly reduce subsequent trading activities. Therefore, their losses are limited. On the other hand, winners increase there following trading intensity to further exploit their above average trading ability. As a result, we should expect to see a positive skewness and persistence in performance. The empirical results here imply that mutual fund industry is more like a repeated game illustrated in Mahani and Bernhardt (2007) rather than a single shot decision model I present in previous sections.

Table 3.2. Comparing Excess Return Statistics for Index and Growth Funds (1994–2004)

This table shows the standard deviation, the skewness, and the kurtosis for excess returns of our mutual fund data over different time frequency. We compare these statistics for our 10856 fund-years of index funds and the 36657 fund-years of growth funds during the period from 1994 to 2004. The difference is marked as bold if it is statistically significant at 5% level after 1000 times of bootstrapping.

Stats	Standard Deviation			Skewness			Kurtosis		
Freq:	Index Funds	Growth Funds	Mean Diff	Index Funds	Growth Funds	Mean Diff	Index Funds	Growth Funds	Mean Diff
Jan	0.0255	0.0292	-0.0038	0.1285	0.7842	-0.6209	7.7160	26.9122	-18.4909
Feb	0.0426	0.0552	-0.0127	2.7896	2.9401	-0.1419	23.3220	22.1652	1.1484
Mar	0.0301	0.0378	-0.0078	-2.1204	-2.1085	-0.0152	14.0729	14.0598	0.0322
Apr	0.0311	0.0362	-0.0051	0.5593	-0.1295	0.7034	6.2504	10.0693	-3.7297
May	0.0244	0.0280	-0.0036	0.5588	-0.1094	0.6652	6.1086	12.8209	-6.9934
Jun	0.0275	0.0330	-0.0055	1.5615	1.8794	-0.3447	16.2079	21.2140	-5.0197
July	0.0247	0.0299	-0.0051	-0.6499	-0.3143	-0.3239	5.5797	6.3434	-0.7793
Aug	0.0253	0.0298	-0.0045	0.4330	0.4521	-0.0149	6.0467	12.3996	-5.9671
Sept	0.0289	0.0335	-0.0049	-3.8409	-0.5046	-3.0784	114.3649	27.2497	60.5325
Oct	0.0303	0.0340	-0.0036	-0.2756	-0.4814	0.2171	6.6431	9.9180	-3.1520
Nov	0.0311	0.0390	-0.0079	-1.5223	-0.4037	-0.9727	40.5737	21.2663	16.9658
Dec	0.0322	0.0369	-0.0048	1.5577	1.4648	0.1309	12.7441	13.0610	-0.0021
Q1	0.0536	0.0601	-0.0063	0.1410	0.3501	-0.2112	7.7915	13.6490	-5.9391
Q2	0.0468	0.0537	-0.0070	0.8386	0.6364	0.1604	5.7494	11.9585	-6.0771
Q3	0.0445	0.0576	-0.0133	-1.3591	0.6165	-2.0034	27.7334	27.4010	-0.0366
Q4	0.0659	0.0885	-0.0227	1.8729	1.8947	0.0259	26.7839	26.2096	0.9431
Annu.	0.1411	0.1584	-0.0172	1.6907	1.8264	-0.1484	23.2616	21.5198	2.2870

3.6.3 Risk Change Within a Year

According to my model as well as Carpenter's (2000), a fund manager tends to increase the risk when an evaluation point is imminent (e.g., the end of year), especially when she is under-performing at that time. Chevalier and Ellison (1997) find that funds alter their portfolios to take more risk in the last quarter of the year. This means that funds should have higher return variance near the end of the year. In Figure 3.3, I find that in our 11 years of samples, 4th quarter return variance is more than 20% higher than those for the other three quarters. However, a close look of the return variance of each year in Figure 3.4 reveals that the abnormally high excess return variance in 4th quarter is mainly caused by the extraordinarily high variance in years 1999 and 2000, which makes the result more likely to be caused by the limitation in data length.

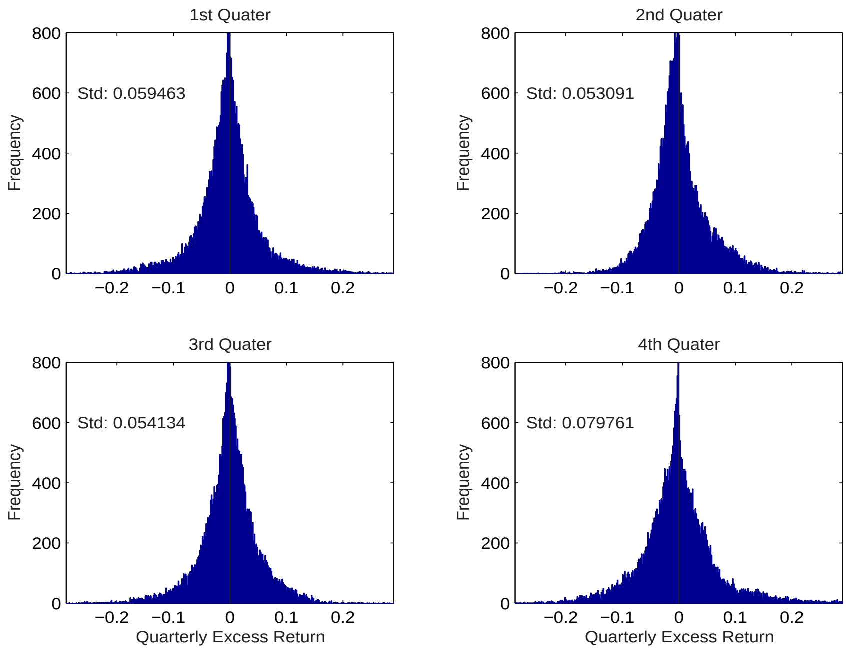

Figure 3.3. Return Variances in Four Quarters of the Calendar Year using Pooled data during 1994-2004.

I then look at the monthly return variances. There is no significant difference towards the end of the year. February has a higher variance but this is caused by

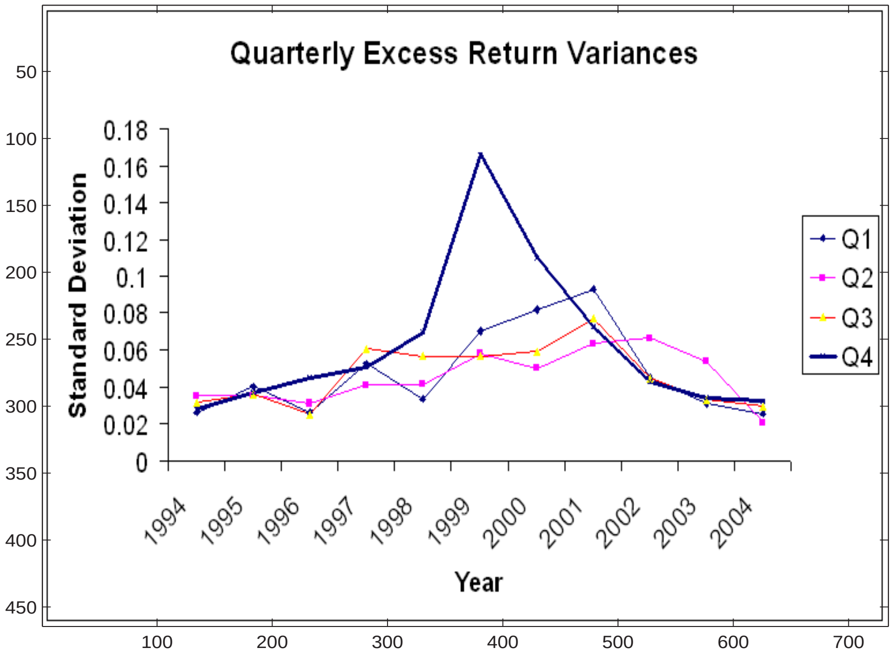

Figure 3.4. Quarterly Excess Return Variance During 1994-2004

outliers as we can see from Figure 3.5.

3.6.4 Risk Changes in Deciles

One implication of the model is that for managers slightly below an expectation or a benchmark, they are more likely to take on additional risks to beat the target. To test this, I categorize all funds into 10 performance deciles with decile 1 to be the lowest return and decile 10 to be the best performers. Then, I compare the annual risks and excess returns across last year's performance deciles as shown in Figure 3.6.

In addition to performance persistence as documented in Carhart (1997), I find those funds in the last year's 3rd and 4th deciles take more risks and beat the performance of the 5th decile. This is consistent with our theoretical predictions. Also, funds in the 2nd, 3rd and 4th deciles take higher risks than funds in the 7th, 8th and 9th deciles, indicating that below average funds generally take more risks than those already above the threshold. Interestingly, those funds in the 5th

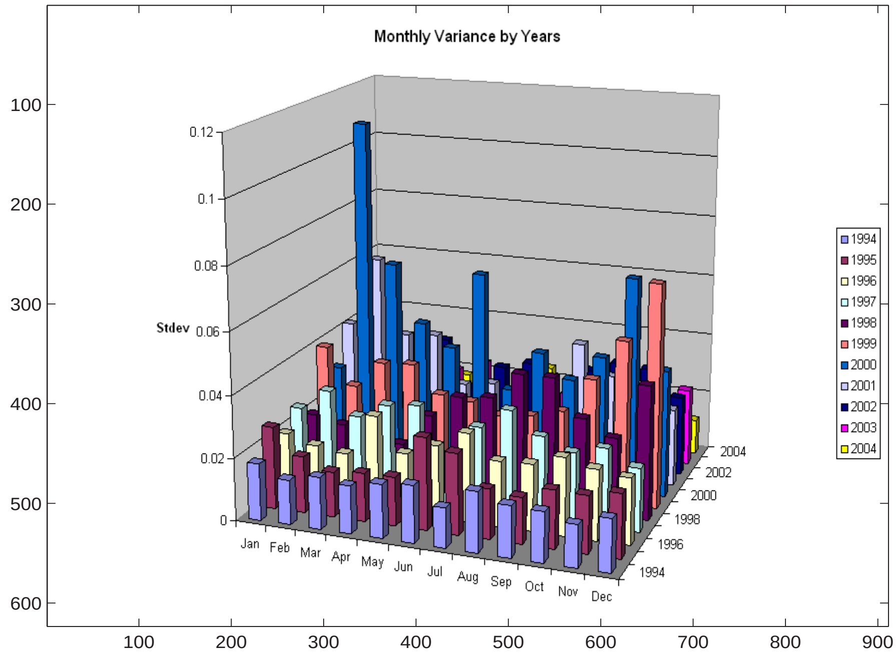

Figure 3.5. Monthly Excess Return Variance During 1994-2004

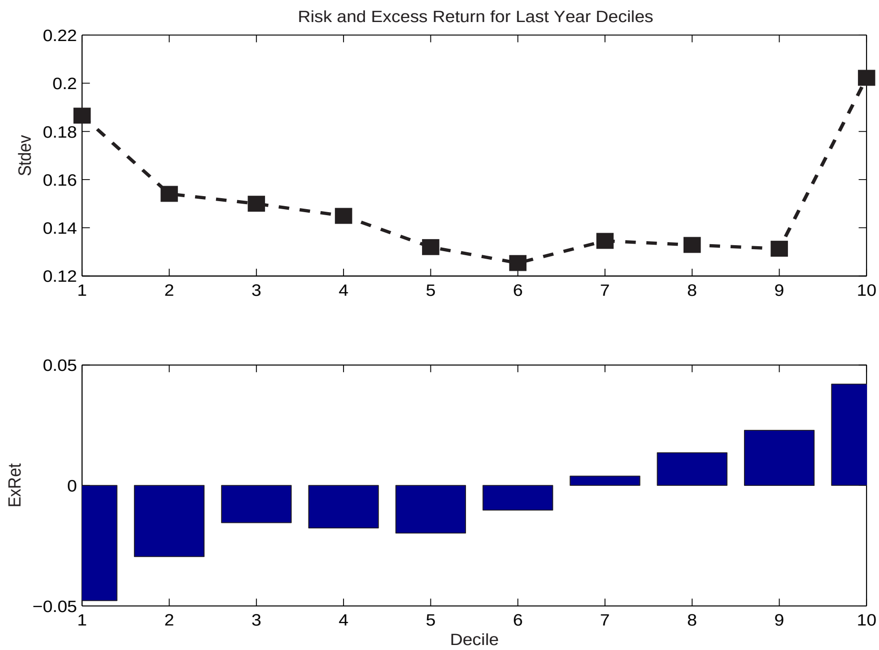

Figure 3.6. Risk and Excess Return by Deciles.
Stdev is the current standard deviation of the annual excess returns for funds ranked within the same decile last year. ExRet is the current average excess return for that decile.

and 6th deciles have the lowest return variance. These funds are in a precarious position that the managers may just want to reduce the variance to lock-in their recent ranks, as my model predicted.

I also find that funds with very low rank (deciles 1 and 2) and very high rank (decile 10) take higher risks. For the high variances of badly performing funds, the models in Grinblatt and Titman (1989) and Carpenter (2000) are more pertinent. For managers perform badly, their payoff function is more like an option rather than a linear reward in our model. Therefore, they increase the risk level to take advantage of the option-like payoff. For funds doing extremely well, there could be a self selection bias as found by Elton, Gruber and Blake (2003) that funds with incentive fees earn positive excess return, but at the cost of bearing more risks. Therefore, those extremely good funds may be intrinsically prone to high risks.

The decile partitioning may not be fine enough to reflect the true risk patterns in mutual fund. Figure 3.7 illustrates the result when funds are partitioned into 20 duo-deciles. The previous findings are robust under this partition.

3.7 Conclusion

I solve a model for the manager's optimal choice problem when there is an option to take a risky project. The theoretical result suggests an explanation that the terminal value of a firm can have a negative-skewed distribution without earnings management. My model also adds to the mutual fund literature by showing that manager's inefficient risk decisions could arise without option-like incentive fees. Empirically, I test the model using the US growth mutual fund data. Funds with slightly below average performance have high excess return variance and on average perform better in the next year than average funds. However, the return distribution is positively skewed, implying that successful managers may repeatedly trade with their superior ability as modeled in Mahani and Bernhardt

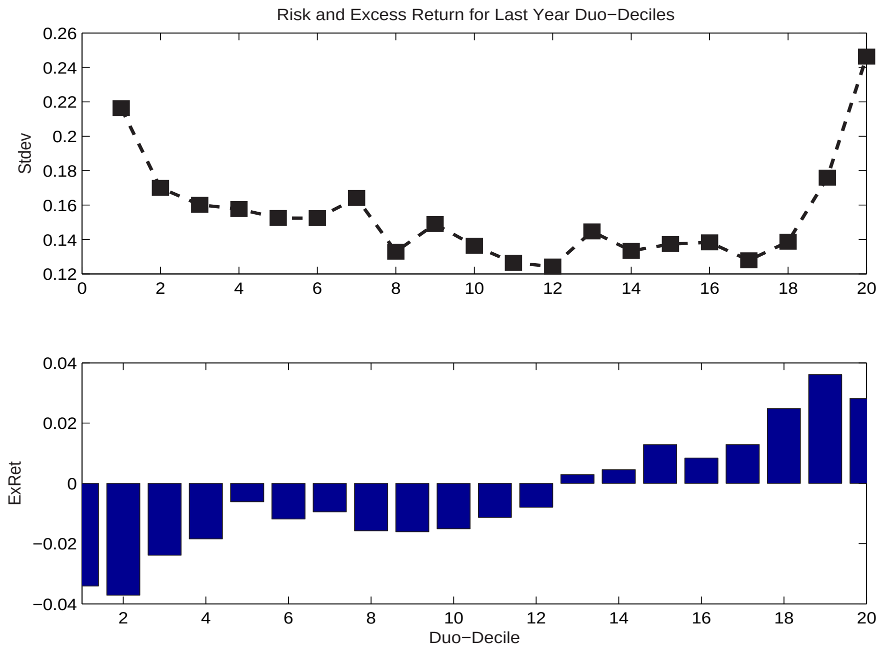

Figure 3.7. Risk and Excess Return by duo-deciles.
Stdev is the current standard deviation of the annual excess returns for funds ranked within the same duo-decile last year. ExRet is the current average excess return for that duo-decile.

(2007).

In our model, the cost of having a threshold or benchmark is the suboptimal risk choice made by the manager. However, such monitoring activity will increase the value for investors by reducing potential agency problem. A possible future research topic could be studying the frequency of monitoring and the firm value. There must exist an equilibrium if: 1) More frequent monitoring will increase manager's effort and in turn increase investor's value. 2) Monitoring is costly because managers take suboptimal portfolio choices such as described in this paper.

Given the empirical findings here and in Chevelier and Ellison (1997) that funds increase their risk exposure near year end, a possible research extension could be adding one more time period into the model and setting a terminal value variance even without taking the risky project. By doing so, we can study the effect of time in the manager's risky decision. The intuition is that when there is plenty of time before the evaluation date, the manager does not have much pressure to take risks. As the evaluation comes imminent, she is more likely to do something if she still falls behind. This could be interpreted in the sense of option pricing as in Carpenter (2000), where an out-of-the-money option has a high price when there is a long time before expiration while it is almost worthless right before the expiration date.

Another simplification left in this research is the fixed level of risk for the risky project. Having a continuum of different projects with different risks will make the model more realistic and yield more implications. However, by doing so we will no longer have the existence condition for a threshold in proposition 1, that is, the manager can always take some levels of risk for the firm if she likes. Whether this is close to or away from reality relies on the empirical findings about whether or not managers abstain from profitable projects just because of the lack of risk choices.

3.8 Appendix

3.8.1 Proof of Proposition 1

A threshold m^* makes the manager indifferent of whether taking or not taking the project: $E[U|e = 0, m = m^*] = E[U|e = 1, m = m^*]$. Therefore, m^* is the solution, if any, of the following equation:

$$U(m^*) = U(m^* - \frac{1}{2}\alpha S) + Pr(m^* + x \geq v)U(B) \tag{3.5}$$

Since $y^* = m^* + x$ is a normal random variable with mean m^* and variance S, the probability in the right hand side is: $1 - \Phi\left(\frac{v-m^*}{\sqrt{S}}\right) = \Phi\left(\frac{m^*-v}{\sqrt{S}}\right)$. Plugging in the CARA utility function in equation 3.1, we can get the equation for m^* in Proposition 1. Let us rewrite it here:

$$\frac{1}{U(B)}\left(e^{-\alpha m^*}\right)\left(e^{(1/2)\alpha^2 S} - 1\right) = \Phi\left(\frac{m^* - v}{\sqrt{S}}\right) \tag{3.6}$$

The left hand side (LHS) of equation 3.6 monotonically decreases with respect of m^* while the right hand side (RHS) increases. As $m^* \to 0$, $LHS \to \frac{1}{U(B)}\left(e^{(1/2)\alpha^2 S} - 1\right) > 0$. Given that v is sufficiently large, $RHS \to 0$ as $m^* \to 0$. On the other boundary, when $m^* = v$, $LHS = \frac{1}{U(B)}\left(e^{-\alpha v}\right)\left(e^{(1/2)\alpha^2 S} - 1\right)$ and $RHS = 1/2$. A unique threshold m^* exists if and only if $LHS|_{m^*=v} < RHS|_{m^*=v}$. This completes the proof.

QED.

3.8.2 Proof of Corollary 1

m^* decreases as $U(B)$ and B increase is immediate from equation 3.3. To find $\frac{dm^*}{dS}$ and $\frac{dm^*}{d\alpha}$, let

$$F = \frac{1}{U(B)}\left(e^{-\alpha m^*}\right)\left(e^{(1/2)\alpha^2 S} - 1\right) - \Phi\left(\frac{m^* - v}{\sqrt{S}}\right) = 0 \tag{3.7}$$

By implicit function theorem:

$$
\begin{aligned}
\frac{dm^*}{dS} &= -\frac{(\partial F/\partial S)}{(\partial F/\partial m^*)} \\
&= \frac{[1/(2U(B))]\alpha^2 e^{-\alpha m^*} e^{(\alpha^2/2)S} - (v-m^*)/(2S^{3/2})\phi(\frac{m^*-v}{\sqrt{S}})}{(\alpha/U(B))e^{-\alpha m^*}(e^{(1/2)\alpha^2 S}-1) + (1/\sqrt{S})\phi(\frac{m^*-v}{\sqrt{S}})} \qquad (3.8) \\
\frac{dm^*}{d\alpha} &= -\frac{(\partial F/\partial \alpha)}{(\partial F/\partial m^*)} \\
&= \frac{(\partial F/\partial \alpha)}{(\alpha/U(B))\,[U'(m^*) - U'(m^* - (1/2)\alpha S)] - (1/\sqrt{S})\phi(\frac{m^*-v}{\sqrt{S}})} \qquad (3.9)
\end{aligned}
$$

The sign of equation (3.8) is undetermined. While the denominator in equation (3.9) is positive, the sign of $(\partial F/\partial \alpha)$ is undecided also, because:

$$
\begin{aligned}
\frac{\partial F}{\partial \alpha} &= \frac{\partial \left\{[U(m^*) - U(m^* - (1/2)\alpha S)]\,/U(B)\right\}}{\partial \alpha} \\
&= [m^* U'(m^*) + (S\alpha - m^*)U'(m^* - (1/2)\alpha S)]\,/U(B) \\
&\quad - [U(m^*) - U(m^* - (1/2)\alpha S)] \cdot B \cdot U'(B)/[U(B)^2] \qquad (3.10)
\end{aligned}
$$

QED.

3.8.3 Proof of Proposition 2

Since the transformation from m to y under the existence of $m^* < v$ is:

$$
y = \begin{cases} m & \text{if } y \leq m^* \text{ or } y \geq v \\ m + x & \text{if } m^* < y < v \end{cases} \qquad (3.11)
$$

the density function of y is a convolution of two normal density functions of the initial firm value and the risky project: $f(m)$ and $g(x)$ in the risk taking range $[m^*, v]$, plus the density of the initial firm value if it is outside the range of taking risk.

$$h(y) = I((y \leq m^*) or (y \geq v)) f(y) + \int_{m^*}^{v} f(y-x) g(x) dx \tag{3.12}$$

Replacing the density functions $f(y-x) = \phi\left(\frac{y-x-m}{\sqrt{S_m}}\right)$ and $g(x) = \phi\left(\frac{x}{\sqrt{S}}\right)$ yields the result.

QED.

Chapter 4

Time-inconsistent Risk Preferences in a Laboratory Experiment (Co-Authored with Jeremy Ko)

4.1 Introduction

There is a substantial literature in psychology and economics that studies time-inconsistent discounting or preferences for consumption over time. The basic conclusion of this work is that people generally prefer immediate to delayed consumption but that this preference reverses when this same decision is translated into the future.[1] This paper explores time-inconsistency in the other main dimension of investment preferences, i.e., risk-preferences. Specifically, we seek to address the question of whether people's future preferences for risk are consistent with their present preferences and to characterize the manner in which they may be inconsistent.

We study this issue with a simple experiment where subjects play a multi-

[1]For overviews of the literature on time-inconsistent discounting, see Ainslie (1992), Loewenstein and Prelec (1992), and Rubinstein (2003).

period betting game. In the experiment, we asked subjects to plan out all of their contingent betting decisions in an initial session (the plan session) then had them actually play the game dynamically in a second session (the play session) one week later to see if these decisions matched their initial plan. We were able to articulate a few simple findings. First, the majority of subjects in our experiment were time-inconsistent; specifically, 72% of subjects bet differently than they had planned in their actual game play. Second, subjects on average took more risk than they had planned in their initial bet as well as after a loss where this increase was more significant when the loss was more recent. Finally, the increase in risk-taking is associated with an increase in the degree of break-even mental accounting, i.e., the desire to return to one's original level of wealth.

Our results are related to documented examples of temporal inconsistency related to lack of self-control with regard to addictive behaviors or impulsive behaviors (Gruber and Koszegi (2001), Read and van Leeuwen (1998)), procrastination (Ariely and Wertenbroch (2002)), and effort (Fang and Silverman (2007) and Paserman (2007)), which are at odds with prior or subsequent desires. Our paper is also related to studies which find that people can not maintain their commitments to prior preferences. For example, several studies have found that people systematically fail to maintain prior commitments to health club attendance (Della Vigna and Malmendier (2006)) and switching credit cards after the introductory rate expires (Shui and Ausubel (2004)). Behavior in our play session reflects similar loss of self-control with regard to impulsive gambling behavior and unwillingness to maintain prior commitments.

The neuroeconomic study of McClure, et al. (2004) is particularly illuminating with regard to temporal inconsistency and our results. These authors examine neural activity when people are presented with immediate versus delayed riskless intertemporal choices. Their main finding is that the brain's limbic system, which is associated with impulsive behaviors, is preferentially activated for immediate rewards but not delayed ones. This neural pattern is likely to be driving the observed behavior in our own experiment as well. Namely, the immediacy of the game during the play session causes subjects to be driven by emotional impulses such as excitement and regret, which causes them to bet more and chase their previous wealth levels more aggressively than they had planned. Our findings are significant because they shed light on the conditions under which emotions exacerbate suboptimal behaviors such as excessive risk-taking and by-products of mental accounting like the disposition effect documented in investment behavior by Odean (1998) and others.

The remainder of this paper is organized as follows. Section 4.2 describes the design of our experiment, and Section 4.3 discusses our major results. Section 4.4 concludes.

4.2 Experimental Design

Our experiment consisted of two sessions, the plan session and the play session, each conducted one week apart. In both sessions, subjects made decisions for a three-round betting game where they could make fair double-or-nothing bets on the outcomes of the spin of a wheel, which was numbered from 1 to 100. In each of the three rounds, they decided how much to bet with the constraint that their bet not exceed their available funds. They could also decide whether to bet on odd or even numbers on the wheel. Subjects were initially given 100 units or $10 for the first bet the game, and available funds in any subsequent round were simply the previous funds plus (minus) the profit (loss) from the previous bet.

In the first session, the plan session, subjects made all of their contingent betting decisions in advance. In other words, they chose their initial bet, their bets for the second round contingent on winning and losing the first bet, and their bets for the third round for every possible contingency of winning or losing the first two bets. Subjects, therefore, made seven betting decisions (i.e., how much to bet and on which set of numbers) for every node of the contingency tree for the three bets. We label these bets: bet 1 (the first round bet), bet 2w and bet 2l (the second round bets contingent on winning and losing the first bet, respectively), bet 3ww and bet 3ll (the third round bets contingent on winning and losing both prior bets, respectively), bet 3wl (the third round bet contingent on winning the first bet and losing the second bet). and bet 3lw (the third round bet contingent on losing the first bet and winning the second bet).

We also asked subjects their beliefs about their probability of winning to determine if their inferences were rational. All data for our experiment were collected by computer using the experimental economics program, Z-tree. A sample screenshot for the plan decision is shown in Figure 4.1. We instructed subjects in the plan session that they were making decisions for a game whose outcomes would be determined in a week. They were told that they would observe the outcomes from spinning the wheel and answer questions related to the experiment in the following week's session.[2]

In the second session, the play session, subjects actually played the game dynamically. Specifically, they decided on their initial bet, and the wheel was then spun and outcomes determined for this bet. The same procedure was then followed for the second bet and then for the third bet. Subjects viewed their decisions from the plan session, but were instructed that they were not required to adhere to this plan and were free to alter these decisions in any way. In order to prevent social influences from affecting preferences, communication between subjects was prohibited in both the plan and play sessions. As in the plan session,

[2]There were only two subjects in the experiment who asked us privately whether they would be allowed to change their betting decisions in the next session. We did not want decision-making to be affected by a feeling of being deceived or cheated in the experiment. We, therefore, told these individuals that that they may or may not be allowed to change their decisions depending on the occurrence of certain events, but that they should take their plan decisions seriously and think of them as final.

we polled subjects in the play session for their chance of winning each bet. A sample screenshot for the play decision is shown in Figure 4.2.

Our subjects consisted of 214 undergraduate finance majors recruited from a finance course at Penn State University. These students were compensated with their earnings from the experiment (the initial $10 plus (minus) any profits (losses) from the three rounds of betting) in addition to a fixed amount of extra credit for their participation in both sessions.

4.3 Results

The goal of this paper is to study whether the risk-taking of subjects in the game was time-consistent and to characterize any time-inconsistency observed. Our main interest, therefore, is in studying the change in risk-taking between the plan and play sessions at various nodes of the game's contingency tree. To this end, we employ three different measures to quantify the risk characteristics of bets.

The first measure is simply the dollar amount of the bet, and the second is its portfolio weight or the proportion of wealth bet. Our third measure captures whether or not subjects may be engaging in "break-even" mental accounting as in Thaler and Johnson (1990) and trying to return to their original wealth of $10. This measure is an indicator variable equal to one if their maximum wealth in the next period is not less than $10 and zero otherwise. We should add that this measure does not definitely determine whether changes in betting behavior between the plan and play are caused by changes in mental accounting. For example, a fixed dollar increase in bets between plan and play could increase the frequency of breakeven bets despite the fact that subjects' degree of mental accounting does not change between sessions. In order to control for the endogeneity of changes in dollar and breakeven betting, one would need to instrument using subjects' wealth in the game which should have a differential effect on dollar bets than on break-even betting. This multivariate analysis requires more data than is available in our experiment to be done meaningfully. We can, therefore, only address whether changes in risk are associated with changes in breakeven betting as an indication of whether mental accounting may have changed between sessions.

There are several reasons why differences in plan and play betting behavior may not reflect inconsistency between present and future risk-preferences on the part of subjects. One reason is that subjects may dynamically update their probabilities of outcomes during the game in the play session. Namely, subjects may improperly believe that the wheel is not fair and that past outcomes influence the odds of subsequent outcomes. This updating would not be reflected in our course game tree from the plan session, which only has win and loss contingencies and not individual numerical outcomes. To control for this possibility, we eliminated all subjects from our data who reported any probability of winning not equal to 50% in any of their plan or play responses.

Another possibility is that subjects' preferences changed over the week between the plan and play sessions for rational reasons. An unexpected change in a subject's economic circumstances over the course of the week such as receiving a parking ticket, for example, might change his or her betting preferences in the game. For this reason, we asked our subjects in the exit interview whether they had experienced an unexpected change in economic circumstances over the past week that may have changed their betting behavior in the game. We eliminated all subjects from the data who answered yes to this question. After applying these filters, we were left with exactly half or 107 of 214 subjects in our sample.

4.3.1 Average Bet Characteristics

We start in table 1 by reporting average bet characteristics for our three measures and average wealth at each node of the game tree. In our data, there is clearly some evidence of Thaler and Johnson's (1990) house money effect, whereby people take more risk after gains because they view betting these additional funds as "playing with the house's money." Namely, the average bet 3ww was greater in dollar terms than bet 2w which was greater than bet 1 for both the plan and play samples. The observation that subjects had a greater propensity to bet after wins is slightly muddied by the budget constraint in our game, i.e., subjects had more available funds to bet after gains. It is unlikely that this constraint played a major role, however, since less than 10% for bets 1 and 2w and roughly 15% for bet 3ww for both sessions were constrained by their budget and bet all of their available funds. In addition, the fact that the budget constraint was more binding in the highest wealth node (bet 3ww) serves to weaken the appearance of the house money effect in our data since more subjects at this node would, in principle, like to bet more than their budget. This fact strengthens our argument for its existence.

One can also see from table 1 that there were general increases in bet amounts and mental accounting between plan and play. We study this issue more carefully in the next sections by computing average bet changes across subjects while attempting to control for wealth differences between plan and play.

4.3.2 Average Bet Changes

77 of 107 subjects (72%) in our sample were time-inconsistent. Namely, the amount that these subjects bet in their actual play was different than in their plan at some point in the game. 33 of our 107 subjects changed their initial bet (bet 1) when they played the game, and the average change was significantly riskier.

In the next two sections, we discuss average changes in the second and third bets of the game. Changes in these second and third round bets, however, may not reflect time inconsistency in these bets if the subject alters his or her first bet. If subjects change their first bet, their wealth in the play session is different than planned at later nodes. Such subjects may change these later bets not because they improperly anticipate their preferences for these bets but rather because their preferences are dependent on the level or path of wealth in the game. In order to control for the prior path of wealth, we compute bet changes between plan and play at each node for two types of samples. In the "no prior change" samples, we compute the average bet change at each node for all subjects who have an outcome at that node in the play session and that have not changed any of their prior bets. Hence, their paths of wealth in plan and play are identical up to that point. We also compute changes at each node for all subjects in the total sample with an outcome at that node in order to increase the size of the sample and to characterize the behavior of subjects that have changed decisions at prior nodes.

- No Prior Change Samples

  Results for the no prior change samples are shown in table 2. In these samples, subjects bet significantly more than they had planned both in terms of dollars and portfolio weight immediately after a loss for bets 2l and 3wl. The average change in risk for bet 3ll is not statistically significant though we lack statistical power in this observation since there were only 8 subjects at this node who had not changed their prior bets. Only one subject in this sample changed this final bet, which became riskier than planned.

  This increase in risk-taking immediately after a loss is associated with an increase in the degree of break-even mental accounting. By our mental accounting measure, subjects on average took significantly more break-even bets than they had planned for bets 2l and 3wl.

- Total Sample

  In the total sample, subjects again took significantly more risk than they had planned on average in terms of their dollar bet immediately after a prior loss at the bet 2l and bet 3wl nodes as can be seen in table 3. The average increase in dollar risk for bet 3ll is again not statistically significant. In the total sample, this lack of statistical significance is probably due to the fact that subjects had less wealth on average when they actually played the game than in their plans for this bet (i.e., they bet more than planned and lose twice) as seen in table 1 and faced a budget constraint.

Subjects did, however, increase their risk-taking in bet 3ll in terms of portfolio weight. On average, subjects bet a greater proportion of their portfolio than they had planned at all nodes with a prior loss (i.e., bet 2l, bet 3wl, bet 3lw, and bet 3ll) and not just immediately after a loss. Subjects exhibited significantly more break-even mental accounting than they had planned for bet 3wl but not for bet 2l in contrast to our no prior change sample. Both sets of results are again explained by fact that most subjects had lower wealth in the play than the plan session for bets 2l, 3lw, and 3ll. Namely, subjects in the total sample bet a greater proportion of their total wealth in the play than in the plan session for bets 3lw and 3ll. In addition, there was a lack of significant increase in mental accounting for bet 2l and bet 3ll because many subjects had wealth below $5 in the play session and could not make breakeven bets.

The greater significance of increases in risk in the whole sample as well as the no prior change samples at the bet 3wl node relative to the bet 3lw node probably reflects a short-term emotional impulse to increase risk and chase prior levels of wealth immediately after a loss. One can see further evidence in table 3 that the average play bet was greater in dollar terms for bet 3wl than for bet 3lw in the play session while the average wealth at these nodes were comparable as seen in table 1.

In summary, we were able to articulate a few simple findings in our analysis. First, the preponderance of subjects in our experiment were time-inconsistent and altered their bets when they played the game in spite of being able to see their planned betting strategy. Second, subjects bet more than they had planned in their initial bet and after a loss. This increase in risk-taking was more significant for changes in the dollar bet and mental accounting and less significant for changes in portfolio weight in the no prior change than in the total sample because of differences in wealth for the latter between plan and play. In addition, this increase was more significant for more recent than more distant losses. Although we can not say definitively that these increased risks after losses were caused by increased mental accounting, the associated increase in the breakeven betting leads us to believe that subjects may be more impulsively chasing prior levels of wealth in the play session. As mentioned previously, these results are consistent with the findings of McClure, et al. (2004) that people are more intensely driven by impulses such as greed and regret when faced with immediate payoffs.

4.4 Conclusion

The findings in this paper are important because they tell us that immediacy between decision-making and outcomes may exacerbate behaviors such as excessive risk-taking and by-products of mental accounting documented in investment behavior such as the disposition effect. One implication of our research is that the level of detachment from outcomes in different decision frames will affect investment. For example, decisions made through a broker are likely to be quite different than decisions made through a financial planner for reasons of framing alone. This line of research also has potential implications for market regulation and design. For example, policymakers have recently become concerned with individuals displaying behavioral biases in the self-management of privatized social security accounts. Our study points at ways in which to mitigate certain biases. One possible mechanism would be to create a delay between decision-making and implementation as a way of dulling impulsive behaviors on the part of accountholders.

Our paper opens the door for interesting paths of future research. One natural question would be whether this research has any implications for pricing in asset markets. An important aspect of our experiment is that it documents a fundamental violation of the Arrow-Debreu theorem. Namely, the static complete market of our plan session was not equivalent to the dynamically complete market of our play session. It would be interesting to study whether our findings have any implications for pricing and anomalies that can be confirmed in market data much as the literature on time-inconsistent discounting has attempted to develop implications for markets and asset prices.[3]

[3]Luttmer and Mariotti (2003), for example, study a model of asset prices in an exchange economy with time-inconsistent agents.

Figure 4.1. Screenshot for Plan Session

Figure 4.2. Screenshot for Play Session

Table 4.1. Average Bet Characteristics

Panel A lists the average bet characteristics and wealth at each node for the plan session whereas Panel B lists these figures for the play session. Bet 1, bet 2w, bet 2l, bet 3ww, bet 3wl, bet 3lw and bet 3ll are subsets including all players in that node of the game's event tree with available data. We summarize the average wealth, bet in terms of dollar amount, portfolio weight, and break-even mental accounting for each node of the game's event tree. The values for wealth and dollar bet are in the game unit of 10 cents. The values for mental accounting are 1 if the maximum wealth in the subsequent period (dollar bet + current wealth) is greater than or equal to the initial wealth of 100 and 0 otherwise. Subjects with zero wealth at a given node receive a null value for our portfolio weight measure.

Panel A:			Plan Session		
Subset:	# of Obs.	Wealth	Dollar Bet	Portfolio Weight	Mental Accounting
Bet 1	107	100	39.5514	0.39551	1
Bet 2w	107	139.5514	63.8692	0.42947	1
Bet 2l	107	60.4486	24.1121	0.45784	0.15888
Bet 3ww	107	203.4206	111.0093	0.4846	1
Bet 3wl	107	75.6822	36.5981	0.52796	0.65421
Bet 3lw	107	84.5607	42.6262	0.53267	0.74766
Bet 3ll	107	36.3364	20.8411	0.65823	0.1028
Panel B:			Play Session		
Subset:	# of Obs.	Wealth	Dollar Bet	Portfolio Weight	Mental Accounting
Bet 1	107	100	44.9439	0.44944	1
Bet 2w	62	144.8387	63.0161	0.42194	1
Bet 2l	45	54.9111	28.4889	0.59103	0.42222
Bet 3ww	32	209.2188	108.3125	0.49033	1
Bet 3wl	30	77.7333	62.5667	0.87138	0.96667
Bet 3lw	25	77.72	53.52	0.7077	0.76
Bet 3ll	20	31.7	23.4	0.85102	0.25

Table 4.2. Average Bet Changes for the No Prior Change Samples

This table presents average changes in bet characteristics across subjects for the "no prior change" samples, i.e. subjects at each node who have not changed any of their prior bets. We present the number of total observations at each node (#Obs), average play bet (Play), average plan bet (Plan), and the mean change (Change) in bet between plan and play at the following nodes (subsets): bet 1, bet 2w, bet 2l, bet 3ww, bet 3wl, bet 3lw and bet 3ll are subsets with an outcome at that node in their game play and who have not changed any of their prior bets. We measure the change in three different measures: dollar amount in Panel A, portfolio weight in Panel B, and degree of break-even mental accounting in Panel C. We have eliminated subjects with zero wealth from our portfolio weight samples. The statistical significance of the mean change is computed under the paired t-test.

Panel A:			Dollar Amount		
Subset:	# of Obs.	Play	Plan	Change	P-Value
Bet 1	107	44.9439	39.5514	5.3925***	0.000662
Bet 2w	45	62.1556	65.9111	-3.7556	0.14273
Bet 2l	29	27	21.6552	5.3448**	0.01763
Bet 3ww	20	119.45	122.45	-3	0.80413
Bet 3wl	10	56.5	32.6	23.9**	0.026541
Bet 3lw	12	42.9167	35.4167	7.5	0.47663
Bet 3ll	8	32.25	30.375	1.875	0.35062
Panel B:			Portfolio Weight		
Subset:	# of Obs.	Play	Plan	Change	P-Value
Bet 1	107	0.44944	0.39551	0.053925***	0.000662
Bet 2w	45	0.42443	0.44516	-0.02027	0.23196
Bet 2l	29	0.54092	0.45868	0.0822**	0.025634
Bet 3ww	20	0.52835	0.51229	0.015254	0.71794
Bet 3wl	10	0.84242	0.54436	0.3525**	0.024536
Bet 3lw	9	0.67016	0.57016	0.1	0.4999
Bet 3ll	8	0.87075	0.79932	0.0714	0.35592
Panel C:			Mental Accounting		
Subset:	# of Obs.	Play	Plan	Change	P-Value
Bet 1	107	1	1	0	NaN
Bet 2w	45	1	1	0	NaN
Bet 2l	29	0.48276	0.37931	0.10345*	0.083061
Bet 3ww	20	1	1	0	NaN
Bet 3wl	10	1	0.7	0.3*	0.081126
Bet 3lw	12	0.66667	0.66667	0	NaN
Bet 3ll	8	0.375	0.375	0	NaN

***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Table 4.3. Average Bet Changes for the Total Sample

This table presents our basic time-inconsistency results for the total sample. We present the number of total observations (#Obs), play bet (Play), average plan bet (Plan), and the mean change (Change) in bet between plan and play at the following nodes (subsets): bet 1, bet 2w, bet 2l, bet 3ww, bet 3wl, bet 3lw and bet 3ll are subsets including subjects with an outcome at that node in their game play. We measure the change in three different measures: dollar amount in Panel A, portfolio weight in Panel B, and degree of break-even mental accounting in Panel C. We have eliminated subjects with zero wealth from our portfolio weight samples. The statistical significance of the difference between play and plan is computed under the paired t-test.

Panel A:	Dollar Amount				
Subset:	# of Obs.	Play	Plan	Change	P-Value
Bet 1	107	44.9439	39.5514	5.3925***	0.000662
Bet 2w	62	63.0161	61.0645	1.9516	0.62223
Bet 2l	45	28.4889	23.0222	5.4667**	0.027179
Bet 3ww	32	108.3125	97.6875	10.625	0.33673
Bet 3wl	30	62.5667	36.0667	26.5***	7.18E+00
Bet 3lw	25	53.52	39.96	13.56	0.15148
Bet 3ll	20	23.4	20.65	2.75	0.40569
Panel B:	Portfolio Weight				
Subset:	# of Obs.	Play	Plan	Change	P-Value
Bet 1	107	0.44944	0.39551	0.053925***	0.000662
Bet 2w	62	0.42194	0.42234	-0.00039	80.98568
Bet 2l	45	0.59103	0.46762	0.1234***	0.000667
Bet 3ww	32	0.49033	0.42614	0.062182	0.1628
Bet 3wl	30	0.87138	0.58297	0.31766***	0.000339
Bet 3lw	21	0.7077	0.50785	0.19985**	0.018586
Bet 3ll	20	0.85102	0.638	0.2130**	0.035832
Panel C:	Mental Accounting				
Subset:	# of Obs.	Play	Plan	Change	P-Value
Bet 1	107	1	1	0	NaN
Bet 2w	62	1	1	0	NaN
Bet 2l	45	0.42222	0.42222	0	1
Bet 3ww	32	1	1	0	NaN
Bet 3wl	30	0.96667	0.76667	0.2**	0.011655
Bet 3lw	25	0.76	0.8	-0.04	0.57431
Bet 3ll	20	0.25	0.15	0.1	0.16255

***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.

Chapter 5

Future Research

The first essay proposes a very strong assessment about market efficiency. Although the performance of each individual anomaly decreases significantly after its publication, an investor can still beat the market simply by picking the best performer among all published anomalies. Therefore, this finding raises a question to every paper on asset-pricing anomalies, including this one itself: how well will the results of published research hold in the future? In other words, what is the overall impact of academic publication on empirical findings? Fama and French (2000) find that the profitability of a firm follows a mean reverting process; if one industry earns an above average return, competition will eventually drive its profitability back to the equilibrium level. Would this be the case for the profitability of asset-pricing anomalies after their publication?

One immediate future research topic could be the study of profitability patterns for asset-pricing anomalies over time. As illustrated in Table 2.7 already, there are no significant time trends on either the relative or the absolute performance of asset-pricing anomalies other than the initial drop after publication. However, this initial result misses some important aspects of real-world investment. For example, profitability may be related to the overall market environment. Excess returns in good years may not be as valuable as excess returns (or reduced losses) in bad years. Also, the overall profitability of all market participants in a particular asset-pricing anomaly is related to the participation pattern over time. If profitability follows mean reversion and investors chase performance as assumed in Chapter 1, the overall profitability for all participants needs to be reevaluated to get a more accurate judgement about market efficiency.

Bibliography

Ainslie, G., 1992. Picoeconomics. Cambridge University Press, New Jersey.

Alexander, G. J., 2000. On back-testing "zero-investment" strategies. Journal of Business 73 (2), 255–278.

Ariely, D., Wertenbroch, K., 2002. Procrastination, deadlines, and performance: Self-control by precommitment. Psychological Science 13 (3), 219–224.

Banz, R. W., 1981. The relationship between return and market value of common stocks. Journal of Financial Economics 9 (1), 3–18.

Baquero, G., Verbeek, M., 2007. Do sophisticated investors believe in the law of small numbers? Working paper.

Barras, L., 2007. International conditional asset allocation under real time uncertainty. Unpublished working Paper, HEC, University of Geneva and FAME, Geneva, Switzerland.

Basu, S., 1977. Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis. Journal of Finance 32 (3), 663–682.

Berk, J. B., Green, R. C., 2004. Mutual fund flows and performance in rational markets. Journal of Political Economy 112 (6), 1269–1295.

Bhandari, L. C., 1988. Debt/equity ratio and expected common stock returns: Empirical evidence. Journal of Finance 43 (2), 507–528.

Bhardwaj, B. K., Brooks, L. D., 1992. The january anomaly: Effects of low share price, transaction costs, and bid-ask bias. Journal of Finance 47 (2), 553–575.

Black, F., 1972. Capital market equilibrium with restricted borrowing. Journal of Business 45 (3), 444–455.

Black, F., Jensen, M. C., Scholes, M., 1972. The capital asset pricing model: Some empirical tests. in: Michael C. Jensen, Studies in the Theory of Capital Markets, Praeger Publishers Inc.

Bondt, W. F. M. D., Thaler, R., 1985. Does the stock market overreact? Journal of Finance 40 (3), 793–805.

Bossaerts, P., Hillion, P., 1999. Implementing statistical criteria to select return forecasting models: What do we learn? Review of Financial Studies 12 (2), 405–428.

Breen, W., Glosten, L. R., Jagannathan, R., 1989. Economic significance of predictable variations in stock index returns. Journal of Finance 44 (5), 1177–1189.

Brennan, M. J., Xia, Y., 2001. Assessing asset pricing anomalies. Review of Financial Studies 14 (4), 905–942.

Brown, K. C., Harlow, W., Starks, L. T., 1996. Of tournaments and temptations: An analysis of managerial incentives in the mutual fund industry. The Journal of Finance 51 (1), 85–110.

Burgstahler, D., Dichev, I., 1997. Earnings management to avoid earnings decreases and losses. Journal of Accounting and Economics 24, 99–126.

Carhart, M. M., 1997. On persistence in mutual fund performance. The Journal of Finance 52 (1), 57–82.

Carpenter, J. N., 2000. Does option compensation increase managerial risk appetite? The Journal of Finance 55 (5), 2311–2331.

Chan, L. K. C., Hamao, Y., Lakonishok, J., 1991. Fundamentals and stock returns in japan. Journal of Finance 46 (5), 1739–1764.

Chevalier, J., Ellison, G., 1997. Risk taking by mutual funds as a response to incentives. Journal of Political Economy 105 (6), 1167–1200.

Cohen, R. B., Polk, C., Vuolteenaho, T., 2002. Does risk or mispricing explain the cross-section of stock prices? Working paper, Harvard University and Northwestern University, Cambridge, MA, and Evanston, IL.

Cooper, M., 1999. Filter rules based on price and volume in individual security overreaction. Review of Financial Studies 12 (4), 901–935.

Cooper, M., Gulen, H., 2006. Is time-series-based predictability evident in real time? Journal of Business 79 (3), 1263–1292.

Cooper, M., Gutierrez, R., Marcum, B., 2005. On the predictability of stock returns in real time. Journal of Business 78 (2), 469–499.

Cooper, M. J., McConnell, J. J., Ovtchinnikov, A. V., 2006. The other january effect. Journal of Financial Economics 82 (2), 315–341.

Coval, J. D., Shumway, T., 2005. Do behavioral biases affect prices? The Journal of Finance 60 (1), 1–34.

Cremers, K. M., 2002. Stock return predictability: A bayesian model selection perspective. The Review of Financial Studies 15 (4), 1223–1249.

Das, S. R., Sundaram, R. K., 2002. Fee speech: Signaling, risk-sharing, and the impact of fee structures on investor welfare. The Review of Financial Studies 15 (5), 1465–1497.

Degeorge, F., Patel, J., Zeckhauser, R., 1999. Earnings management to exceed thresholds. The Journal of Business 72, 1–33.

DellaVigna, S., Malmendier, U., 2006. Paying not to go to the gym. American Economic Review 96 (3), 694–719.

Elton, E. J., Gruber, M. J., Blake, C. R., 2003. Incentive fees and mutual funds. The Journal of Finance 58 (2), 779–804.

Elton, E. J., Gruber, M. J., Das, S., Hlavka, M., 1993. Efficiency with costly information: A reinterpretation of evidence from managed portfolios. The Review of Financial Studies 6 (1), 1–22.

Fama, E. F., 1970. Efficient capital markets: A review of theory and empirical work. Journal of Finance 25 (2), 383–410.

Fama, E. F., 1990. Stock returns, expected returns, and real activity. Journal of Finance 45 (4), 1089–1108.

Fama, E. F., French, K. R., 1988. Dividend yield and expected stock returns. Journal of Financial Economics 22 (1), 3–25.

Fama, E. F., French, K. R., 1992. The cross-section of expected stock returns. Journal of Finance 47 (2), 427–465.

Fama, E. F., French, K. R., 2000. Forecasting profitability and earnings. Journal of Business 73 (2), 161–175.

Fang, H., Silverman, D., 2007. Time-inconsistency and welfare program participation: Evidence from the nlsy. Yale University Working Paper.

French, K. R., 1980. Stock returns and the weekend effect. Journal of Financial Economics 8 (1), 55–69.

French, K. R., 2008. The cost of active investing. Working paper, Dartmouth College, Hanover, NH.

Goyal, A., Welch, I., 2003. Predicting the equity premium with dividend ratios. Management Science 49 (5), 639–654.

Goyal, A., Welch, I., 2007. A comprehensive look at the empirical performance of equity premium prediction. Unpublished working Paper, Emory University and Yale University.

Grinblatt, M., Titman, S., 1989. Adverse risk incentives and the design of performance-based contracts. Management Science 35 (7), 807–822.

Grossman, S. J., Stiglitz, J. E., 1980. On the impossibility of informationally efficient markets. American Economic Review 70, 393–408.

Gruber, J., Koszegi, B., 2001. Is addiction "rational"? theory and evidence. Quarterly Journal of Economics 116 (4), 1261–1303.

Gruber, M. J., 1996. Another puzzle: The growth in actively managed mutual funds. The Journal of Finance 51 (3), 783–810.

Guttman, I., Kadan, O., Kandel, E., 2003. Adding the noise: A theory of compensation-driven earnings management. Working Paper.

Handa, P., Tiwari, A., 2006. Does stock return predictability imply improved asset allocation and performance? evidence from the u.s. stock market (1954-2002). Journal of Business 79 (5), 2423–2468.

Jacklin, C. J., Kleidon, A. W., Pfleiderer, P., 1992. Underestimation of portfolio insurance and the crash of october 1987. Review of Financial Studies 5 (1), 35–63.

Jegadeesh, N., 1990. Evidence of predictable behavior of security returns. Journal of Finance 45 (3), 881–898.

Jin, L., 2005. How does investor short-termism affect mutual fund manager short-termism. Working Paper, Harvard Business School.

Keim, D. B., 1983. Size-related anomalies and stock return seasonality: Further empirical evidence. Journal of Financial Economics 12 (1), 13–32.

Kim, C.-W., Park, J., 1994. Holiday effects and stock returns: Further evidence. Journal of Financial and Quantitative Analysis 29 (1), 145–157.

Kirchenheiter, M., Melumad, N. D., 2002. Can "big bath" and earnings smoothing co-exist as equilibrium financial reporting strategies? Journal of Accounting Research 40, 761–796.

Lakonishok, J., Shleifer, A., Vishny, R. W., 1994. Contrarian investment, extrapolation, and risk. Journal of Finance 49 (5), 1541–1578.

Lee, C. M. C., Swaminathan, B., 2000. Price momentum and trading volume. Journal of Finance 55 (5), 2017–2069.

Lo, A. W., MacKinlay, C., 1990. Data-snooping biases in tests of financial asset pricing models. Review of Financial Studies 3 (3), 431–467.

Loewenstein, G., Prelec, D., 1992. Anomalies in intertemporal choice: Evidence and an interpretation. Quarterly Journal of Economics 107 (2), 573–597.

Mahani, R. S., Bernhardt, D., 2007. Financial speculators' underperformance: Learning, self-selection, and endogenous liquidity. Journal of Finance 62 (3), 1313–1340.

McClure, S. M., Laibson, D. I., Loewenstein, G., Cohen, J. D., 2004. Separate neural systems value immediate and delayed monetary rewards. Science 306 (5695), 503–507.

Naik, N. Y., Ramadorai, T., Stromqvist, M., 2007. Capacity constraints and hedge fund strategy returns. European Financial Management 13 (2), 239–256.

Odean, T., 1998. Are investors reluctant to realize their losses? Journal of Finance 53 (5), 1775–1798.

Parrino, R., Weisbach, M., 1999. Measuring investment distortions arising from stockholder-bondholder conflicts. Journal of Financial Economics 53 (1), 3–42.

Paserman, M. D., 2007. Job search and hyperbolic discounting: Structural estimation and policy implications. Hebrew University Jerusalem Manuscript.

Penman, S. H., 1987. The distribution of earnings news over time and seasonalities in aggregate stock returns. Journal of Financial Economics 18 (2), 199–228.

Pesaran, H., Timmermann, A., 1995. Predictability of stock returns: Robustness and economic significance. Journal of Finance 50 (4), 1201–1228.

Read, D., van Leeuwen, B., 1998. Predicting hunger: The effects of appetite and delay on choice. Organizational Behavior and Human Decision Processes 76 (2), 189–205.

Rozeff, M. S., 1984. Dividend yields are equity risk premiums. Journal of Porffolio Management Fall, 68–75.

Rozeff, M. S., Kinney, W. R., 1976. Capital market seasonality: The case of stock returns. Journal of Financial Economics 3 (4), 379–402.

Rubinstein, A., 2003. "economics and psychology"? the case of hyperbolic discounting. International Economic Review 44 (4), 1207–1216.

Rubinstein, M., 2001. Rational markets: Yes or no? The affirmative case. Financial Analysts Journal 57 (3), 15–29.

Schwert, G. W., 2003. Anomalies and market efficiency. In: Constantinides, G., Harris, M., Stulz, R. (Eds.), Handbook of the Economics of Finance. Elsevier Science Ltd., North-Holland, Amsterdam 1B, 939–974 (Chapter 15).

Shleifer, A., Vishny, R. W., 1997. The limits of arbitrage. Journal of Finance 52 (1), 35–55.

Shui, H., Ausubel, L. M., 2004. Time inconsistency in the credit card market. University of Maryland Working Paper.

Simin, T. T., Forthcoming. The (poor) predictive performance of asset pricing models. Journal of Financial and Quantitative Analysis.

Sirri, E. R., Tufano, P., 1998. Costly search and mutual fund flows. Journal of Finance 53 (5), 1589–1622.

Solnik, B., 1993. The performance of international asset allocation strategies using conditioning information. Journal of Empirical Finance 1, 33–55.

Stattman, D., 1980. Book values and stock returns. The Chicago MBA: A Journal of Selected Papers 4, 25–45.

Stoll, H. R., 1993. Equity trading costs in-the-large. Journal of Portfolio Management 19 (4), 41–50.

Stoll, H. R., Whaley, R. E., 1983. Transaction costs and the small firm effect. Journal of Financial Economics 12 (1), 57–79.

Thaler, R., 1985. Mental accounting and consumer choice. Management Science 4 (3), 199–214.

Thaler, R. H., Johnson, E. J., 1990. Gambling with the house money and trying to break even: The effects of prior outcomes on risky choice. Management Science 36 (6), 643–660.

White, H., 2000. A reality check for data snooping. Econometrica 68 (5), 1097–1126.

Zeelenberg, M., 1999. Anticipated regret, expected feedback and behavioral decision making. Journal of Behavioral Decision Making 12, 93–106.

Zhu, Y., Zhou, G., 2007. Technical analysis and theory of finance. Working paper.

Vita

Zhijian Huang

Education

Ph.D. Candidate in Finance, Smeal College of Business, Penn State University, August, 2008

M.Eng. Financial Engineering, Cornell University, May 2003

M.S. Electrical and Computer Engineering, Michigan State University, Aug 2002

B.Eng. Telecommunication Engineering (Major) and Industrial Management (Minor), Shanghai Jiaotong University, July 1998

Publications

"Arrogance can be a Virtue: Overconfidence, Information Acquisition, and Market Efficiency." (with Jeremy Ko), Journal of Financial Economics, Volume 84, Issue 2, May 2007, Pages 529-560

Working Papers

Real-Time Profitability of Published Anomalies: An Out-of-Sample Test

Managerial Decisions and Mutual Fund Performance

Time-Inconsistent Risk Preferences in a Laboratory Experiment (with Jeremy Ko)

Academic Experience

Research Assistant, Department of Finance, The Pennsylvania State University, 2003-2008

Instructor, Financial Markets and Institutions, Department of Finance, The Pennsylvania State University, 2005-2008