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

Zhijian (James) Huang *

September 7, 2007

Abstract

Previous studies show mixed results about the out-of-sample performance of various asset-pricing anomalies. To remove data-snooping bias, this paper simulates a real-time trader who chooses among all asset-pricing anomalies published prior to that time using only backward-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 considering more anomalies. 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.

JEL Classification: G11, G14, D83

Key Words: Data-snooping Bias; Asset-pricing Anomalies; Out-of-sample Test; Real-time Simulation

---

*The author is from Smeal College of Business, Pennsylvania State University. For this project, I receive many helpful insights from talks with my dissertation committee members: Jean Helwege, Jay Huang, Jeremy Ko, Runze Li, and Tim Simin. I also want to thank Laura Field, David Haushalter, Hong Li and seminar participants at Pennsylvania State University for their comments. All errors are my own. Please address correspondence to: 363A Business Building, University Park, PA 16802, zuh102@psu.edu (e-mail), Phone: 814-865-9201, Fax: 814-865-3362.

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] A typical out-of-sample test involves the partitioning of the time-series data into two parts: a training period before an evaluation point and a testing period after the evaluation point. The result of an out-of-sample test is then determined by the performance in the testing period using a model derived from only the training period data. The evaluation point can be an arbitrary time point in the time series or simply the time when the first draft of a paper is written. In a more elaborate setting, the evaluation point can also change as a time point rolling forward in a step-ahead out-of-sample test. In all cases, however, the evaluation point is set at a time before the anomaly is published. This means, at the hypothetical evaluation time in an out-of-sample test, there were not many real world investors aware of the anomaly of interest. In contrast, after being published, an anomaly is well known and widely exploited by many traders. This difference in public attention before and after publication could affect the future profitability of a published anomaly, even if the anomaly has sustained an out-of-sample test using historical data before its publication. Namely, investor attention and participation may adversely affect the magnitude of an anomaly or a profitable trading strategy. As Schwert (2003) summarizes, many asset-pricing anomalies attenuated or even totally disappeared after initial publication. In the mutual fund and hedge fund literature, it is well documented that fund profitability is decreasing over size (e.g. Berk and Green, 2004; Naik, Ramadorai, and Stromqvist, 2007). Anecdotally, the famous 1987 stock

[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).

market crash is largely blamed for a widely used dynamic hedging strategy called "portfolio insurance", a strategy with its performance sensitive to how many other people are using it simultaneously (e.g., see Jacklin, Kleidon, and Pfleiderer, 1992). Therefore, a research question arises: based *only* on the academically published asset-pricing anomalies and their associated trading strategies, how much excess return, if any, could be realized *ex ante* by a rational trader who efficiently processes information in real time?

In this paper, I run an out-of-sample test in which I assume a real-time trader chooses trading strategies from published anomalies only. In other words, the simulated real-time trader does not possess the ability to identify an asset-pricing anomaly until it is published. Unlike other real-time investment papers, I do not allow a researcher to exogenously determine which anomalies a trader should consider in real time. In stead, I rely on the sophisticated procedure of academic research to unfold anomalies over time. This process guarantees that all profits a simulated trader earned in the past are from published anomalies well known to other market participants at that time, which is exactly the same situation a trader faces if he is about to exploit any published anomalies in future trading. Similar to the real-time investment literature such as Pesaran and Timmermann (1995) and Cooper, Gutierrez, and Marcum (2005), I simulate a trader who progressively picks the best performing anomaly over time out of an anomaly universe. Particularly, in my simulation not all anomalies are available at all times. Instead, only those anomalies that have already been published in a top finance journal enter into the real-time trader's scope. For example, the strategy of "investing in small firms" is assumed not to be considered until 1981, when Banz (1981) first publishes his seminal work about the size anomaly. At any historical point in time the simulated trader can only use the then-available data as well as the then-available technology (i.e. those then-published asset-pricing anomalies in this case) to make investment decisions for the time being.

One important procedure in this paper is to implement a group of anomalies to form an anomaly universe from which the simulated trader forms his real-time investment decisions based on the availability of anomalies. Obviously, it is impossible to encompass every asset-pricing anomaly ever published into this study. To reduce the number of anomalies to a feasible range without incurring forward-looking bias, I apply three filters into the anomaly selection process: anomaly type, publishing journal, and rebalancing frequency. 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 easy to backtest because they only use market return and firm characteristics data that was widely available back to when these anomalies were first published. Second, I assume the real-time trader in my simulation only looks for anomalies in two of the top finance journals: the Journal of Finance (JF) and the Journal of Financial Economics (JFE).[2] I look through every paper in the above two finance journals from 1977 to 1993 looking for anomalies that might have caught the attention of a real-time trader.[3] Although I do get hints from some recent survey papers such as Schwert (2003), I rely on the issue-by-issue search to finalize the true set of strategy universe faced by a real-time trader over time. This search process ensures that anomalies are chosen because they caught a real-time trader's attention, regardless whether or not they are famous nowadays. Finally, to reduce small sample bias, I exclude those long-term return anomalies that require

[2]Another top finance journal, the Review of Financial Study, became available at a relatively late time when most of the calendar and cross-sectional return anomalies had already been published. Therefore, I exclude it in this research.

[3]The searching range for this study may incur some bias because year 1977 and 1992 just cover two important cross-sectional anomalies: the earning/price ratio anomaly (Basu, 1977), and the book-to-market anomaly (Fama and French, 1992). A robustness check is done to check what if we just omit these two anomalies by looking from 1978 to 1991 and find similar results.

a holding period of more than one year to generate excess return.

My major finding is that a trader who picks the best published anomalies based on backtesting result and trade with it can earn considerable returns in excess to a benchmark return. More specifically, I find a real-time trader can significantly beat the market in annual excess return by average margins ranging from 5.64% to 12.63% in different cases. This result is based on annual model re-calibration of picking the best performer among anomalies that were available by then. With a training period of two to five years, the excess return is even higher. This excess return also comes with better statistics in other commonly used performance measures such as the Sharpe ratio and the certainty equivalent rate of returns. Similar results are acquired when the trader only considers calendar anomalies or cross-sectional return anomalies. Also, as I increase the number of strategies available in the real-time trader's total strategy 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.

This paper contributes to the real-time investment literature by running out-of-sample test controlling for not only data snooping bias but also survivorship bias of chosen anomalies. In this simulation, I endogenize the choice of anomalies to certain types of published anomalies in major finance journals. At any evaluation point of the out-of-sample test, this process of choosing anomalies removes the potential influence of subsequent knowledge of anomaly performance. Therefore, this research provides guideline for the practice of applying published anomalies into future trading activities. Results in this paper suggest there is significant excess return relative to the market benchmark if a real-time trader keeps trading with the anomaly that has the best recent performance. This finding also validates the common investment practice of chasing past performance of trading strategies in real time.

Even though I assume the trader chooses from all published anomalies after filters, there could still be a potential of data snooping bias in this research because of the large freedom in the simulation design. However, I believe this bias is minimized in this research for the following reasons. First, in terms of the quantitative measure proposed by White (2000), the nature of a real-time specification search enables a relatively low chance of data snooping bias. According to White (2000), the likelihood of a data snooping bias is proportional to the number of possible research designs. In this research, the choice of anomaly to test is endoginized, which effectively reduces one dimension of variation in research design. Second, 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 ratio of mixed decoy strategies, and different subsets of anomaly universe. Moreover, I focus on the average to the worst case scenarios, rather than the best result, of the simulations run in this study. I allow for the real-time trader to pick a suboptimal parameter relative to what we can find in hindsight. Finally, I test the robustness of my main conclusions using different sample period. Although this paper is already an out-of-sample test on the profitability of asset-pricing anomalies, there is no guarantee the conclusions found in this paper will keep working in the future. To evaluate the possible out-of-sample performances of this research in the future, I try to switch the calendar back to an earlier time point and repeat the whole set of simulations. By doing this, I can see whether or not I would have found the same results had I conducted this research earlier.

The rest of the paper starts with a literature review in Section 2. Section 3 is the research design that describes the out-of-sample testing scheme and the in-sample performance of my basic anomalies. Section 4 reports the out-of-sample results along three lines: the length of training period, the number of strategies under consideration, and the ratio of non-working decoy strategies. Section 5 does some of the robustness checks including an out-of-sample test on this research and a test assuming traders can do academic research themselves. Section 6 concludes the research and discusses some issues regarding the position of this paper in empirical finance research.

2 Literature Review

The idea of only using the then-available technology in out-of-sample tests is mentioned in several real-time investment papers. For example, Pesaran and Timmermann (1995) briefly discuss that a real-time trader can only use the data as well as the technology available at a historical point in time. Cooper, Gutierrez, and Marcum (2005) add the once believed Beta strategy into trader's strategy universe to make the simulation more realistic because the Beta strategy turned out to be no longer work. However, none of the papers has imposed the restriction that a real-time trader could only notice an anomaly after its initial publication.

This paper is closely related to the real-time investment literature that shows mixed results about the out-of-sample profitability of anomalies. On the one hand, real-time adjustment of predictive models is found unprofitable. Researchers report no or very limited out-of-sample excess return over a benchmark index by recursive specification search over time. Cooper, Gutierrez, and Marcum (2005) and Pesaran and Timmermann (1995) show empirically that many useful predictive variables we know in hindsight cannot, in most cases, be exploited *ex ante*. In addition, Lewellen and Shanken (2002) demonstrate theoretically that predictability in asset prices may not be exploited or even perceived by rational investors because investors need to learn the true data generating process in real time. Goyal and Welch (2003) find that because of poor parameter stability, the popular predictive variable, dividend yield, cannot forecast stock returns out-of-sample.[4] On the other hand, there is evidence that dynamically recalibrating forecasting models in real time can indeed significantly improve profitability over a benchmark. For example, Brennan and Xia (2001) study asset-pricing anomalies by considering the learning of model uncertainties. They find an "open-minded" investor who gives prior probability to both the traditional capital asset pricing model (CAPM) and the empirical findings of Fama and French (1993) earns more in a long horizon than if the investor only believes in the CAPM. [5]

While model over- and under-fitting may contribute to the variation of the out-of-sample performance, the exogenously determined predictive model or predictive parameter set could be a major reason for the mixed results. Cooper and Gulen (2004) argue that the choice of predictive variables in an out-of-sample test, along with some other research design parameters, may cause a data-snooping bias. This paper differs from the out-of-sample profitability test literature by endogenizing the choice of models to only published anomalies from a naive anomaly universe based on a backward-looking literature review. [6]

3 Research Design

In this section, first I introduce the basic trading strategies used in this research, and then I provide a detailed description of the recursive out-of-sample simulation. I focus on several key elements of this research design: training and holding periods, strategy selection criteria, transaction costs, and performance measures.

[4]Another paper showing inconclusive evidence of real-time profit is Handa and Tiwari (2006).

[5]Other papers finding dynamic asset allocation profitable include Breen, Glosten, and Jagannathan (1989), Cremers (2002), Solnik (1993), Xia (2001), and Zhu and Zhou (2007).

[6]Several other papers such as Cooper, Gutierrez, and Marcum (2005), Pesaran and Timmermann (1995), and Bossaerts and Hillion (1999), also endogenize model selection, but they do not consider the model availability and the choice of strategy universe benefits from hindsight.

3.1 Data and the Basic Strategies

There are eight basic trading strategies 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 six basic strategies implemented in this paper are cross-sectional anomalies based on six predictive variables: size, book-to-market (B/M) ratio, momentum, earnings/price ratio (E/P), cashflow/price ratio (CF/P), and dividend yield (Div/P). 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 index from the Center for Research in Security Prices (CRSP) as the market benchmark return. For the returns on cross-sectional portfolios, I use the monthly factor risk portfolio returns data from Kenneth French's data library.[7] The risk-free data is from the Fama Risk Free Rates in CRSP. 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 1 lists the initial publication of the anomaly, the data availability range, and a brief description of how the strategy works for each basic strategy.

As we can see from Table 1, most anomalies were published in the 1970's and the 1980's. Note that the publications listed in the second column may not be the earliest publication of

[7] I acknowledge Kenneth French for making these data sets available at his Web site at: http://mba.tuck.dartmouth.edu/ pages/faculty/ken.french/data_library.html.

Table 1: Information about the Basic Anomalies

This table summarizes the basic information about the eight 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. JF is for the Journal of Finance. Panel A lists two strategies based on calendar anomalies. Panel B lists eight cross-sectional trading strategies based on seven 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), and leverage ratio (D/E).

	Panel A: Strategies Based on Calendar Anomalies		
Anomaly Name	Initial Publication	Data Availability	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 Availability	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
Momentum	Jegadeesh (1990), JF	1926-2006	Buy past winners
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

* The anomaly I implement 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 slight 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 through the trading period.

Table 2: The In-Sample Performance of Basic Anomaly

This table demonstrates the in-sample performance of each basic anomalies. *Pub_Year* is the publication year of the anomaly. *Ex_Before* (*Ex_After*) is the average annual excess return over the benchmark return before (after) publishing. *Wealth* is the trader's terminal wealth in dollar amount at the end of year 2006 if he starts with $1 in the publishing year after considering transaction costs. *Wealth_B* is the terminal wealth for the benchmark index. *Sharpe* and *Sharpe_B* are sharpe ratios for the trading strategy and the benchmark index after publishing, respectively. The Sharpe ratio is calculated by deviding the mean excess return over risk free rate by its stardard 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	Pub_Year	Ex_Before	Ex_After	Wealth †	Wealth_B	Sharpe	Sharpe_B
Monday	1980	9.44%***	-0.79%	5.49	28.06	0.49	0.61
JanEfft	1983	13.04%***	9.62%***	87.8	18.03	1.21	0.66
Size	1981	11.49%***	1.35%	15.59	20.9	0.37	0.57
B/M	1991	5.26%**	6.3%*	10.95	5.46	0.79	0.67
Momentum	1990	9.16%***	7.95%*	12.45	5.15	0.63	0.56
E/P	1977	10.67%***	6.09%***	124.28	36.11	0.8	0.57
CF/P	1990	6.58%***	3.76%**	7.85	5.15	0.68	0.56
Div/P	1978	2.89%*	1.32%	38.18	37.72	0.62	0.62

***, **, 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.

the anomaly. Since one filter used in this research is to only include anomalies in two finance journals, there is a possibility that an anomaly may have been appeared in another journal at an earlier time. For example, Stattman (1980) in the journal: "The Chicago MBA: A Journal of Selected Papers" first reports the book-to-market ratio anomaly. However, unlike academics who benefit a lot in hindsight from the growing literature on asset-pricing anomalies, the simulated trader would not know the anomaly until it was first mentioned in Chan, Hamao, and Lakonishok (1991) in the journal of finance.

Table 2 lists the in-sample performance of the basic trading strategies before and after initial publishing. We can see some strategies experienced a decreased anomalous return after being published. The anomalous return is measured by the annual average excess return over a benchmark index. For example, when compared with the benchmark index after its publishing year, the Monday effect has a negative excess return.[8] For other anomalies, they deliver reduced or less significant excess returns after being published. These in-sample results are generally consistent with the 2003 survey article by Schwert (2003).

3.2 The Recursive Trading Scheme

In my simulation, I follow the real-time investment literature to implement a recursive model selection and trading scheme. In addition, I set the initial publication of a strategy as the time point when the strategy becomes publicly available. Below is a narrative description about the big picture of the recursive trading scheme, with key design elements highlighted and discussed in details later.

First, I form an anomaly universe using part or all of the eight basic anomalies. The simulation starts at the earliest publishing year within the anomaly universe, i.e. a trader

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

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 criteria.* After selecting the best anomaly to follow, the real-time trader trades according to that anomaly for the next *holding period.* At the end of the holding period, the trader liquidates his portfolio. The trader repeats the above procedure standing at a time point one *holding period* later using the then-available information. 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 on different *performance measures.* During the trading process, different *transaction costs* are applied to different trading strategies, but there is no transaction cost considered for the benchmark portfolio.

- Training and Holding Periods

  I study the rolling training (or 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 for the one year holding. The reason is that in this simulation the trader starts trading only after those strategies being published in the 1970's or 1980's, leaving less than 30 years for the simulation. Therefore, it seems reasonable to assume a one year holding period and annual portfolio rebalancing.

- Strategy Selection Criteria

  In this research, I use the past excess return as the strategy selection criterion. There are other strategy selection criteria used in the literature. For example, Cooper, Gutierrez, and Marcum (2005) and Pesaran and Timmermann (1995) use the Sharpe ratio and the terminal wealth at the end of training period as alternative model selection criteria and the results are similar. Their criteria, however, cannot be implemented here because of the low frequency annual return data I use. I also do not use any statistical criteria such as those in Bossaerts and Hillion (1999) for model selection because to be consistent with my assumption that the real-time trader learns about strategies based on academic publications, I also assume the trader treats each strategy as a black box without really making judgements based on the structural details of each strategy. This assumption rules out the possibility for the trader to evaluate the fitness of a strategy based on its internal structure, e.g. the number of predictive parameters in the model. Instead, the real-time trader relies purely on the statistics of the past performance to make investment decisions.

- Transaction Costs

Transaction cost in implementing trading strategies is a widely studies topic. Following the "high level" transaction cost in Pesaran and Timmermann (1995), I assume a one percent transaction cost on a round-trip turnover between risk-free T-Bills and stock shares, and a two percent transaction cost on a turnover between different stocks. As we can see from Table 1, except for the Monday Effect, all the basic strategies in this study have an annual turnover ratio of one. Also, among those strategies with one annual turnover, all but the January Effect incur turnovers between the risk-free asset and stocks. Therefore, the transaction cost is two percent annually for the January Effect and one percent 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 a real world investor can hardly neglect it. To solve this dilemma, I assume a five percent 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.

- Performance Measures

To measure the real-time trader's performance, I formally report three performance measures. All three measures are about the performance of the real-time trader relative to the market index. 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 performance is the Sharpe Ratio difference between the simulated return and the market return, measuring the risk-return characteristics of the real-time trader:

$$D\_Sharpe = Sharpe_p - Sharpe_m = \frac{mean(R_p - R_f)}{Std(R_p - R_f)} - \frac{mean(R_m - R_f)}{Std(R_m - R_f)} \tag{1}$$

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.

The third measure is the certainty equivalent rate (CER) of returns used in, e.g. 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 portfolio and the benchmark portfolio as:

$$D\_CER = CER_p - CER_m = E(R_p - R_m) - \frac{1}{2}\alpha[Var(R_p) - Var(R_m)] \tag{2}$$

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

I only calculate statistical significance for the first measure: the annual excess return, based on a standard t test against zero over the time series. For the other two measures, I only calculate an empirical p-Value together with a sample probability density distribution based on multiple simulations run on different subsets of the basic strategies.

Terminal wealth as a performance measure is also calculated but is only reported in some figures in this paper.

- Other Issues

  In this simulation study, I design the trading strategies so as not to allow for short sales.

4 Out-of-Sample Results

In this section, I first set up a basic case for the out-of-sample simulation, then expand the basic case along the variation of three different parameters. The basic case is set with an anomaly universe of the eight basic anomalies under a one year training length. In the following subsections, I show the performance of the real-time trader alone the variations of the training length and the number of anomalies considered.

4.1 Performance with Different Training Lengths

Table 3 shows the simulation result of a trader who picks an available (i.e. published) anomaly based on past performances of various training lengths. I report the real-time trader's returns and the benchmark returns for three different anomaly groups: all anoma-

Table 3: The Out-Of-Sample Performance with Different Training Years

This table presents the out-of-sample performance of a trader who picks the best performing anomaly among published anomalies recursively over time. Panel A through Panel D represent different training lengths of 1 year, 2 years, 5 years, and 10 years. I report three different anomaly groups represented by *Grp_Name*: *All* is for all anomalies; *Calendar* is for the three calendar anomalies only; *Cross* is for the six strategies based on cross-sectional return predictability. For each anomaly group, I report the average annual return: *Rt*, the average annual benchmark return *Bench*, and the standard deviations of them: *Std_Rt* and *Std_Bench*. I also report three performance measures of the real-time return over the benchmark index. *Ex_Rt* is the annual excess return with a statistical significance using a standard *t*-test. I also put in the variance of the excess return *Var_Ex* for reference. *D_Sharpe* is the difference in Sharpe ratios between the real-time trader and the market index. *D_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 *D_Sharpe* and *D_CER* are described in Section 3.2.

	Panel A: 1-Year Training Period							
Grp_Name	Rt	Std_Rt	Bench	Std_Bench	Ex_Rt	Std_Ex	D_Sharpe	D_CER
All	19.3%	21.5%	13.67%	13.98%	5.64%*	16.67%	0.04	2.97%
Calendar	20.24%	16.16%	14.18%	14.15%	6.07%***	10.12%	0.28	5.46%
Cross	19.51%	23.27%	13.67%	13.98%	5.85%*	17.98%	0	2.38%
	Panel B: 2-Year Training Period							
Grp_Name	Rt	Std_Rt	Bench	Std_Bench	Ex_Rt	Std_Ex	D_Sharpe	D_CER
All	24.74%	23.55%	13.67%	13.98%	11.07%***	18.92%	0.21	7.48%
Calendar	21.5%	15.91%	14.18%	14.15%	7.33%***	11.79%	0.36	6.8%
Cross	23.01%	23.03%	13.67%	13.98%	9.35%***	18.22%	0.15	6%
	Panel C: 5-Year Training Period							
Grp_Name	Rt	Std_Rt	Bench	Std_Bench	Ex_Rt	Std_Ex	D_Sharpe	D_CER
All	26.29%	21.1%	13.67%	13.98%	12.63%***	17.23%	0.36	10.13%
Calendar	22.58%	15.71%	14.18%	14.15%	8.4%***	10.81%	0.45	7.94%
Cross	23.59%	22.18%	13.67%	13.98%	9.93%***	16.63%	0.21	6.96%
	Panel D: 10-Year Training Period							
Grp_Name	Rt	Std_Rt	Bench	Std_Bench	Ex_Rt	Std_Ex	D_Sharpe	D_CER
All	22.3%	16.31%	13.67%	13.98%	8.63%***	12.32%	0.41	7.92%
Calendar	21.22%	15.77%	14.18%	14.15%	7.04%***	9.32%	0.36	6.56%
Cross	18.47%	21.51%	13.67%	13.98%	4.81%	16.1%	0	2.14%

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

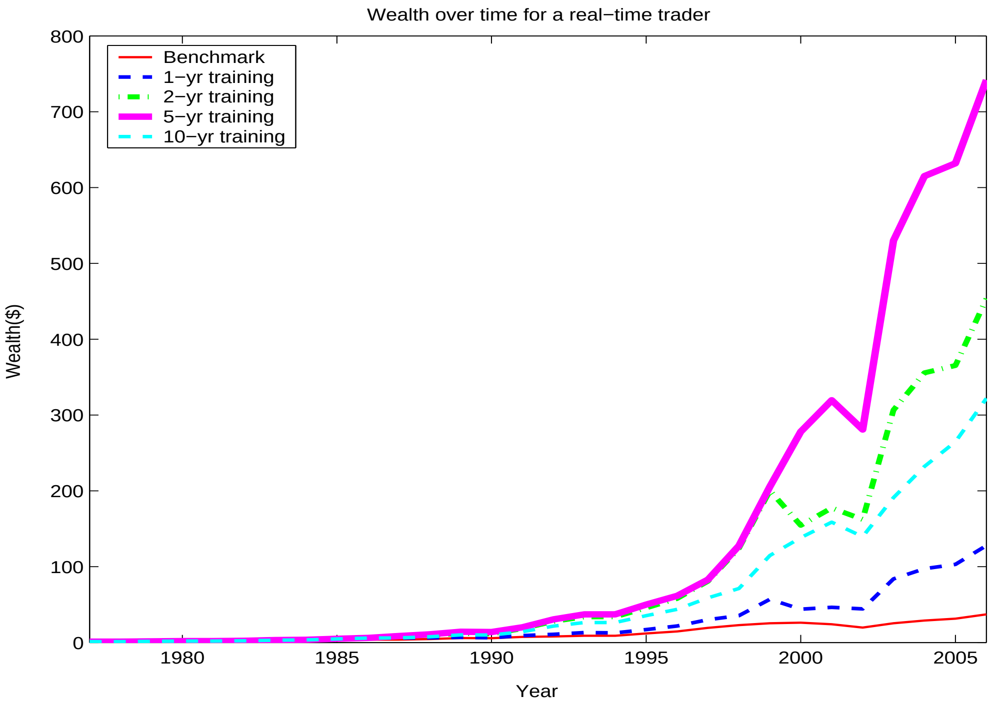

Figure 1: Terminal wealth of the market index and a trader who starting from \$1, picks the best performing strategy over past 1 year, 2 years, 5 years, and 10 years.

lies, calendar anomalies only, and cross-sectional anomalies only. I list three performance measures relative to the benchmark performance: the excess return with significance on a standard t-test, the difference in Sharpe ratios, and the difference in certainty equivalent rate (CER) of returns.

As we can see, in all scenarios in table 3 the real-time trader who chases the performance of published anomalies beats the market as shown in annual excess return (Ex_Rt). The average annual excess return over the buy-and-hold return of a market benchmark ranges from 5% to 13% and are all significant at the 10% confidence level except for one case. 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 for all cases. The simulation from year 1977 (1980 for the calendar anomalies only case) yields a larger Sharpe ratio and a higher CER return than the market return in almost all cases.

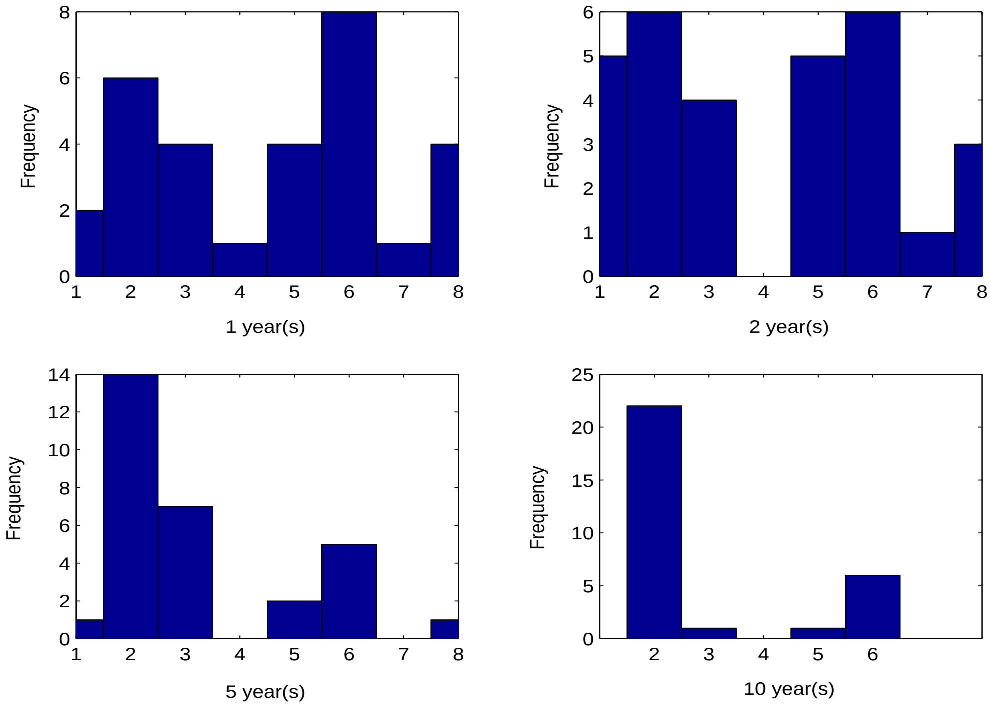

Figure 2: The frequencies of usage when a trader picks from all basic anomalies. The Y-axis is the number of years an anomaly is used. The X-axis represents different anomalies indexed by a sequence number: 1 for the Monday effect, 2 for the January effect, 3 for the size variable, 4 for the B/M variable, 5 for the momentum variable, 6 for the E/P variable, 7 for the CF/P variable, and 8 for the Div/P variable.

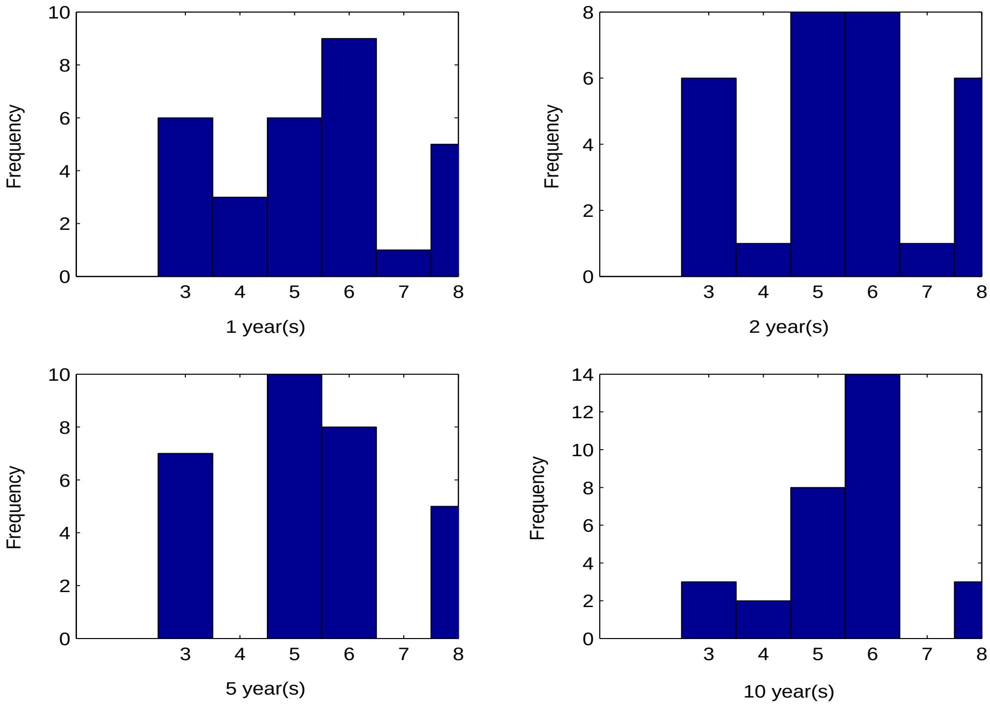

Figure 3: The frequencies of usage when a trader picks from only the cross-sectional anomalies. The Y-axis is the number of years an anomaly is used. The X-axis represents different anomalies indexed by a sequence number: 3 for the size variable, 4 for the B/M variable, 5 for the momentum variable, 6 for the E/P variable, 7 for the CF/P variable, and 8 for the Div/P variable.

While there is no doubt that actively chasing the best published anomaly can yield large excess return over a benchmark, in three out of the twelve cases the simulation even yields an excess return larger than the largest in-sample excess return of a single strategy. (9.62% excess return from the January Effect. See Table 2.)

Figure 1 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 $700 by actively back-testing and trading with published anomalies, 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 4 will show more details about the effect of training length.

Figures 2 and 3 show what anomalies are used by the real-time trader. Figure 2 is for all anomalies and figure 3 is the case when only cross-sectional anomalies are considered by the trader. In figure 2, as the training length increases from one year to ten years, the anomaly selections converge to the January effect which has the best in-sample fitness, yet the 10-year training length does not yield the highest return as we can see from figure 1. In figure 3, although the anomaly selection does not converge to the single best anomaly, the frequencies of usage with 10-year training length roughly reflect the rank of these anomalies in terms of in-sample performance (E.g. the E/P ratio and the momentum are the most frequently used anomalies).

Finally, I aggregate the data in Table 3 plus some unreported cases with other training length to study the relationship between training length and excess return. Figure 4 shows this relationship for all anomalies, calendar anomalies only, and cross-sectional anomalies

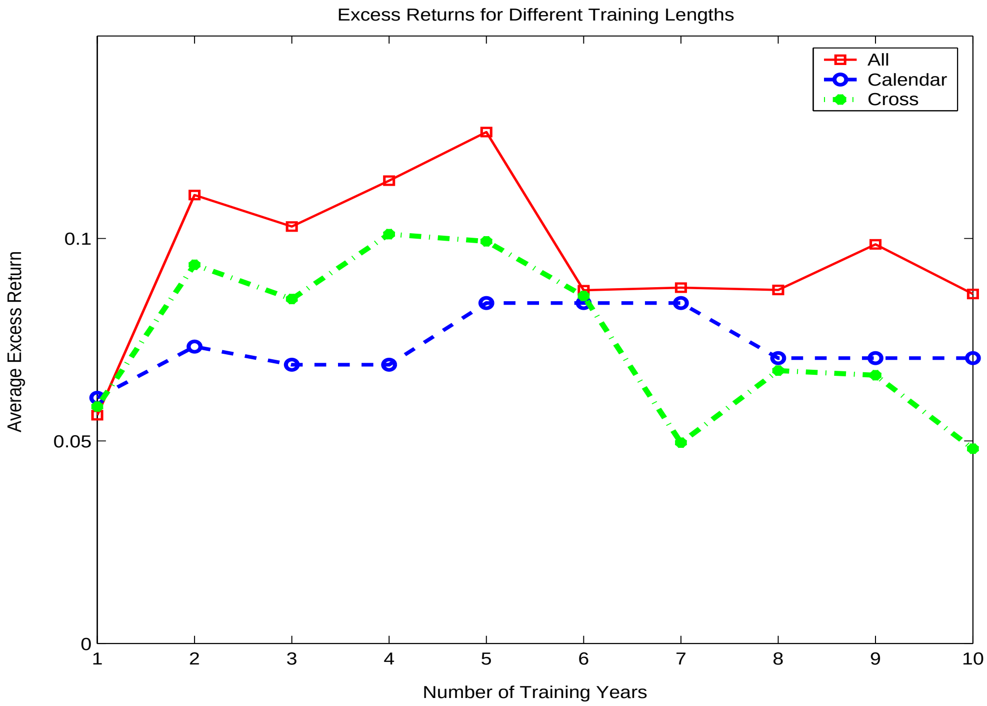

Figure 4: Excess returns over the benchmark index for different training lengths. The X axis represents different training lengths from 1 year to 10 years. The Y axis is the excess return over the benchmark index.

only. As shown in Figure 4, neither a too short (1 year) nor a too long (more than 5 years) training length is optimal for the real-time trader if he picks from all anomalies or from cross-sectional anomalies. In this study, a training period between 2 years and 5 years seems to deliver better out-of-sample performances.

4.2 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 to not to utilize any information from hindsight, it is still an interesting question what if the real-time trader knows more anomalies or fewer anomalies than the eight anomalies implemented in this research. On the

Table 4: The Out-Of-Sample Performance with a Subset of Basic Anomalies

This table presents the out-of-sample performance of a trader who picks the best performing anomaly out of a subset of eight basic anomalies recursively over time. Panel A through Panel C represent different overall anomaly groups. 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, *N_Anomaly*, ranges from one anomaly to the total number of anomalies available within that group. *N_Comb* is the total number of different combinations of anomalies to form a subset of *N_Anomaly* number of anomalies. I report the average annual return *Rt*, the average annual benchmark return *Bench*. I also report three performance measures of the excess return over the benchmark index. *Ex_Rt* is the annual excess return followed by its *P_Value* using a standard t-test. *D_Sharpe* is the difference in Sharpe ratios between the real-time trader and the market index. *D_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 *D_Sharpe* and *D_CER* are described in Section 3.2. The reported values for *Rt*, *Bench*, *Ex_Rt*, *P_Value*, *D_Sharpe*, and *D_CER* are the average values over all cases represented by *N_Comb*.

Panel A: All Anomalies							
N_Anomaly	N_Comb	Rt	Bench	Ex_Rt	P_Value	D_Sharpe	D_CER
1	8	16.24%	13.42%	2.82%	0.2852	0	1.08%
2	28	17.07%	13.73%	3.34%	0.2972	0.03	1.58%
3	56	17.55%	13.89%	3.66%	0.2773	0.02	1.76%
4	70	17.89%	13.91%	3.98%	0.224	0.01	1.88%
5	56	18.21%	13.87%	4.34%	0.1799	0.01	2.04%
6	28	18.53%	13.82%	4.71%	0.1358	0.01	2.25%
7	8	18.89%	13.74%	5.14%	0.1008	0.02	2.56%
8	1	19.3%	13.67%	5.64%	0.0741	0.04	2.97%
Panel B: Calendar Anomalies							
N_Anomaly	N_Comb	Rt	Bench	Ex_Rt	P_Value	D_Sharpe	D_CER
1	2	14.96%	14.05%	0.91%	0.006	-0.01	0.35%
2	1	20.24%	14.18%	6.07%	0.0045	0.28	5.46%
Panel C: Cross-Sectional Anomalies							
N_Anomaly	N_Comb	Rt	Bench	Ex_Rt	P_Value	D_Sharpe	D_CER
1	6	16.66%	13.2%	3.46%	0.3783	0	1.32%
2	15	16.51%	13.51%	3.01%	0.3456	-0.02	0.73%
3	20	16.86%	13.75%	3.11%	0.3477	-0.04	0.63%
4	15	17.43%	13.82%	3.61%	0.2777	-0.05	0.83%
5	6	18.3%	13.77%	4.53%	0.1837	-0.04	1.41%
6	1	19.51%	13.67%	5.85%	0.0853	0	2.38%

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

one hand, due to implementation limit, inevitably there are too few anomalies tested in this paper relative to what is true in real world. On the other hand, the basic anomaly universe in this paper could also be too large relative to a real-time trader's scope. Each trader, in reality, may only focus on a limited number of anomalies, rather than all those published in academic journals. To address this issue, this section and the following section investigate the real-time performance when the actual anomaly universe is larger or smaller than the eight basic anomalies. I shrink the anomaly universe to a subset of the eight basic strategies in this section. In next section, I expand the anomaly universe by mixing extra "decoy" anomalies together with the basic anomalies.

I repeat the simulation in section 4.1 for cases when the trader only knows a subset of the original eight-anomaly universe. This will incur a total of $2^8 - 1 = 255$ different simulations for the all-anomaly case, 3 simulations for the calendar-anomaly-only case, and 63 simulations for the cross-sectional-anomaly-only case. Table 4 reports the performances of these simulations sorted by the number of anomalies in the reduced anomaly universe. As we can see, for the all-anomaly case, the excess return measure (Ex_Rt) and the difference in certainty equivalent return measure (D_CER) are increasing when more anomalies are considered. In the cross-sectional-anomaly case, considering five or six anomalies also generates better performance than considering fewer, in terms of excess return and certainty equivalent return. However, we do not see a monotonic trend in Sharpe ratio. Despite its high returns, chasing past performances at a one-year horizon also brings more return volatility. Table 4 also implies that chasing performances in real time renders a performance that is better than the average performance of all anomalies but worse than the best anomaly. For example, in Panel C of the Table 4 we see the real-time trader can earn an excess returns (4.53% and 5.85%) when considering five and six anomalies. These results are higher than the average in-sample annual excess return of 4.46% (calculated from Table 2) but lower than the best single anomaly performance (7.95% for the Momentum strategy). Overall, results in Table 4 support the value of considering more anomalies and trading with real-time model selection.

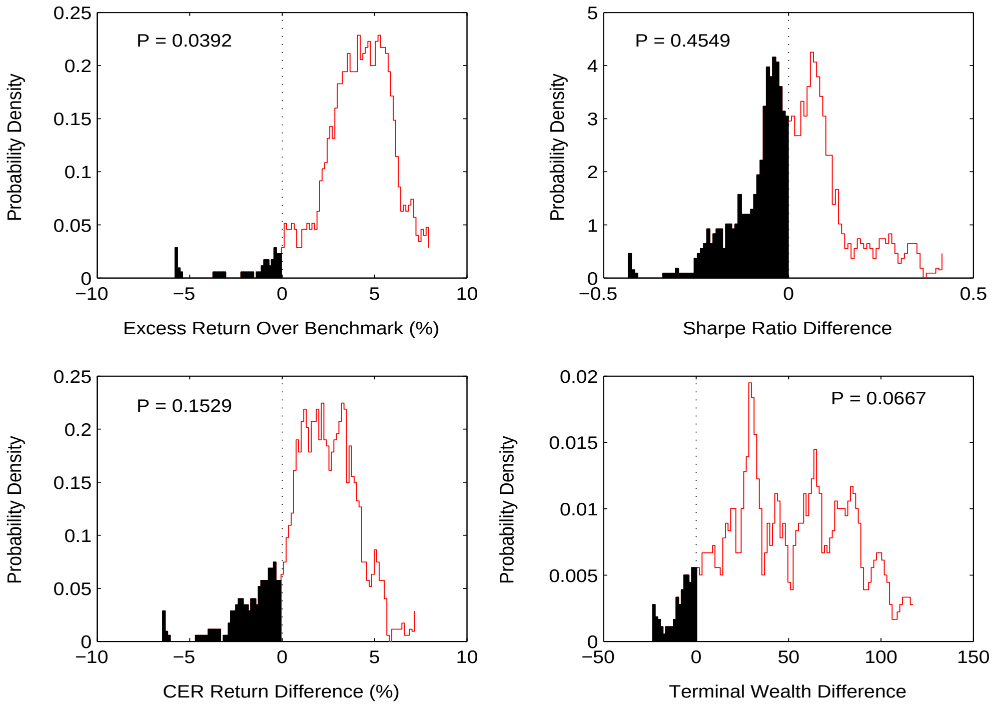

Figure 5: Sample probability density functions (PDFs) for four performance measures: the excess return, the Sharpe ratio difference, the CER return difference, and the terminal wealth difference. These PDFs are estimated using simulations based on the 255 subsets of the eight basic anomalies. Empirical p-Values are reported based on the ratio of negative values among all simulation runs.

The fundamental research question of this paper is: Can a trader beat the market in real time by following published anomalies? In section 4.1 we find positive answer with statistical significance on the excess returns for most cases. 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 strategy subsets. I aggregate the 255 simulation results derived from all subsets of the eight basic anomalies to form empirical probability density functions (PDFs) of the four performance measures: the excess return over benchmark, the Sharpe difference, the certainty equivalent rate of return difference, and the terminal wealth difference. Based on the empirical PDFs, I calculate the empirical p-Values for the four performance measures. Figure 5 lists the PDFs and their corresponding p-Values. As we can see, the excess return measure and the terminal wealth measure generates significantly better performance than the benchmark. However, the other two measures taking into consideration of return volatility do not generate statistically significant result. The difference in measures with and without consideration of risk indicates that most excess returns are generated when the market performs well.

4.3 Performance with Decoy Strategies

In this section, I enlarge the basic anomaly universe by mixing non-working decoy anomalies into it to see if a real-time trader's performance would be affected or not. Similar to Cooper, Gutierrez, and Marcum (2005), I use the non-working deciles of the factor risk portfolios to form a pool of candidate decoy anomalies. For example, "investing in the sixth decile of firm size" is a candidate decoy anomaly. To make sure that decoy anomalies are not too similar to those working anomalies, I exclude the two deciles that are closest to the working decile. For example, the second and the third smallest firm deciles are not decoy anomalies because they are too close to the smallest size decile, but the 4th decile and up are valid candidates. The purpose of adding decoy anomalies is to compensate for the possible bias that anomalies published in top journals are likely to be more effective than those published in other journals.

Table 5 presents the out-of-sample performance with different ratios of mixed decoy anomalies. For example, a decoy ratio of two means in addition to the original eight basic anomalies, another 18 randomly picked non-working anomalies are also considered by the trader in real-time. As we can see from Table 5, for the real-time performance with one-

Table 5: The Out-Of-Sample Performance with Decoy Anomalies

This table presents the out-of-sample performance of a trader who recursively picks over time the best performer out of a pool of eight basic anomalies plus certain ratio of non-working decoy anomalies recursively. Panel A through Panel C reports performance in terms of the excess return, the difference in Sharpe ratios, and the difference in certainty equivalent rate (CER) of returns, respectively and relative to the benchmark. In each panel, I report the performance along with five different decoy ratios and four different training lengths. The decoy ratio is the number of randomly picking decoy anomalies in the number of times of the basic anomalies. A decoy ratio of zero means no decoy anomaly is added to the basic anomaly universe. A significance level of the standard t statistics is reported for the excess return series in Panel A. The detailed calculation of *D_Sharpe* and *D_CER* are described in Section 3.2.

	Panel A: Excess Return			
Decoy Ratio	1 year	2 years	5 years	10 years
0	4.58%*	11.65%***	10.94%***	9.54%***
1	4.24%	11.97%***	10.97%***	10.47%***
2	3.8%	12.35%***	11.8%***	8.96%***
3	3.62%	13.79%***	11.47%***	10.47%***
4	4.6%	12.51%***	11.43%***	9.87%***
	Panel B: Sharpe Ratio Difference			
Decoy Ratio	1 year	2 years	5 years	10 years
0	0.25	0.38	0.39	0.58
1	0.17	0.37	0.39	0.47
2	0.12	0.38	0.37	0.54
3	0.17	0.48	0.4	0.51
4	0.13	0.43	0.37	0.51
	Panel C: CER Return Difference			
Decoy Ratio	1 year	2 years	5 years	10 years
0	4.34%	9.18%	9.08%	9.71%
1	4.15%	10.36%	9.76%	8.79%
2	3.35%	9.33%	9.18%	8.54%
3	1.95%	9.64%	8.85%	10.12%
4	1.71%	10.75%	9.76%	9.85%

***, **, and * indicate statistical significance at the 1%, 5%, and 10% level, respectively.
Note 1: All values in rows with decoy ratio greater than zero are results of a single random simulation and may change in another run.
Note 2: There should be no stars expected in Panel B and C because no statistical significance test is conducted for the Sharpe ratio difference and the CER return difference.

year training period, noisy decoy anomalies do affect on the Sharpe ratio measure and the CER measure. However, the effect on performances with longer training length is trivial. Intuitively, this is because with longer evaluation period, a trader can better distinguish between working and non-working anomalies. Consistent with the findings in section 4.1, a moderate training length is the best to pick out the best performing out-of-sample anomaly without losing new information about the performances of anomalies.

5 Robustness Check

The purpose of this research is to remove data-snooping bias incurred when researchers exogenously choose anomalies to test. Using out-of-sample results that are less subject to data-snooping bias, we can predict future trading profits with more confident. However, this research still falls into the paradigm of empirical asset pricing which relies on historical data to infer a model, and then use the model to predict future returns. Whether or not the conclusions in this research will hold in future still depends on the unknown data that will become available in the future. In this section, I use different data length to check the robustness of the main results of this paper. In order to do this, I repeat the same set of analysis after switching back the calendar to five years ago and ten years ago, and see if the main results still hold. Another real world fact is the lag between when a paper becomes a forthcoming paper and when the paper 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 this, I also repeat the whole set of analysis by switching back the publication year for two years. I expect the results to remain qualitatively the same but with less statistical significance when switching back calendar years, because there are fewer samples in the simulation. But the simulation with earlier publishing year should generate better performances because the real-time trader could take advantage of anomalies at an earlier time.

Therefore, I repeat the real-time simulation for three additional scenarios: with data ending in 2001, with data ending in 1996, and with each publishing year minus two. I do not report all the results here. Instead, I list the following key statistics about the real-time profitability:

- The rank of training lengths

  So far we can see from figure 1 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 rank of training lengths for each robustness check scenario to see if the rank change over time or not.

- The average performances for three performance measures

  In Table 3 I report the out-of-sample performance for three performance measures in 12 cases: four difference training lengths with three different anomaly universes each. I calculate the average performance of these 12 cases for each robustness check scenarios.

- The empirical p-Values

  Figure 5 reports the empirical p-Values for four performance measures. I report these p-Values across all robustness check scenarios

Table 6 compares our main conclusions across four 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, and the forthcoming paper scenario with all publishing years set two years earlier. As we can see, results are similar across these four

Table 6: Main Conclusions under Different Robustness Check Scenarios

This table compares major conclusions about the real-time profitablity in four different scenarios. 2006, 2001, and 1996 represent the situations when the available data ends in year 2006, 2001, and 1996, respectively. "Minus Two" is the case when all the publishing years of anomalies are subtracted by two to compensate for the time as a forthcoming paper. 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_D_Sharpe, and Avg_D_CER are the average values for the three performance measures across the 12 cases presented in Tables 3. The three performance measures are the excess return measure, the Sharpe ratio difference measure, and the CER difference measure. p_Ex_Rt, p_D_Sharpe, p_D_CER, and p_D_Wealth are empirical p-Values for the excess return measure, the Sharpe ratio difference measure, the CER difference measure, and the terminal wealth difference measure, respectively.

Scenario	2006	2001	1996	Minus Two
Rank of Training	5,2,10,1	5,2,10,1	5,2,10,1	5,2,10,1
Avg_Ex_Rt	8.06%	7.76%	5.93%	8.83%
Avg_D_Sharpe	0.24	0.27	0.21	0.24
Avg_D_CER	6.06%	6.36%	5.03%	6.69%
p_Ex_Rt	0.0392	0.102	0.1804	0.0157
p_D_Sharpe	0.4549	0.451	0.4392	0.3137
p_D_CER	0.1529	0.2275	0.2353	0.0706
p_D_Wealth	0.0667	0.1333	0.1961	0.0275

scenarios. 5-year and 2-year training lengths consistently yield better out-of-sample performance than 10-year and 1-year training lengths. Just as we expected, knowing an anomaly early in its forthcoming stage improves real-time trader's performance. In addition to better average performances for the excess return measure and the CER measure, the column "Minus Two" also shows lower p-Values for all performance measures. The significance of findings decreases as we switch calendar back to 2001 and then to 1996 because of the reduced number of samples. However, results remain qualitatively the same as the current scenario with data till 2006. From unreported tables when available data ends in 2001 and in 1996, as well as when the publishing years are two years earlier, I find similar results that the real-time profitability increases as the number of anomalies under consideration increases. Overall, simulation runs using truncated data suggest the main findings in this paper would be the same had I conducted this research project earlier. The alternative assumption that a trader considers an anomaly when it is in forthcoming stage strengthens the real-time performance relative to a benchmark.

6 Conclusion

This paper examines the question whether or not a trader can beat the market by picking the best performing published asset-pricing anomaly. To remove 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 exogenously. My conclusion generally shows that by taking advantage of published anomalies, a trader can beat the market. This result is robust to various choices of several exogenous 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 rely on past performance data to draw inference about future profitability. Also, there are still many possible scenarios in which a real-time trader can under-perform the market. For example, I have not considered trader's learning effect in the sense of Brennan and Xia (2001). At every point in time, the trader devotes himself to the active picking of the best performing anomaly. If we consider the learning effect that a trader may gradually increase his devotion to the active anomaly picking, the result could be weakened.

One way to circumvent the conundrum of predicting the stock market is to soften the question to a more practical one such like "what is the chance of losing the benchmark in the next 10 years?", rather than asking a more direct question of "is the stock market predictable?" or "is the stock market efficient?". Traders with different risk aversion levels and different investment horizons should find the information about past performance useful in answering these more practical questions.

References

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

[2] Guillermo Baquero and Marno Verbeek. Do sophisticated investors believe in the law of small numbers? *Working paper*, 2007.

[3] Laruent Barras. International conditional asset allocation under real time uncertainty. *Unpublished working Paper, HEC, University of Geneva and FAME, Geneva, Switzerland*, 2007.

[4] S. Basu. 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, 1977.

[5] Jonathan B. Berk and Richard C. Green. Mutual fund flows and performance in rational markets. *Journal of Political Economy*, 112(6):1269–1295, 2004.

[6] Laxmi Chand Bhandari. Debt/equity ratio and expected common stock returns: Empirical evidence. *Journal of Finance*, 43(2):507–528, 1988.

[7] Werner F. M. De Bondt and Richard Thaler. Does the stock market overreact? *Journal of Finance*, 40(3):793–805, 1985.

[8] Peter Bossaerts and Pierre Hillion. Implementing statistical criteria to select return forecasting models: What do we learn? *Review of Financial Studies*, 12(2):405–428, 1999.

[9] William Breen, Lawrence R. Glosten, and Ravi Jagannathan. Economic significance of predictable variations in stock index returns. *Journal of Finance*, 44(5):1177–1189, 1989.

[10] Michael J. Brennan and Yihong Xia. Assessing asset pricing anomalies. *Review of Financial Studies*, 14(4):905–942, 2001.

[11] Louis K. C. Chan, Yasushi Hamao, and Josef Lakonishok. Fundamentals and stock returns in japan. *Journal of Finance*, 46(5):1739–1764, 1991.

[12] Michael Cooper and Huseyin Gulen. Is time-series based predictability evident in real time? *Working paper*, 2004.

[13] Michael Cooper, Roberto Gutierrez, and Bill Marcum. On the predictability of stock returns in real time. *Journal of Business*, 78(2):469–499, 2005.

[14] Michael J. Cooper, John J. McConnell, and Alexei V. Ovtchinnikov. The other january effect. *Journal of Financial Economics*, 82(2):315–341, 2006.

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

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

[17] Eugene F. Fama. Stock returns, expected returns, and real activity. *Journal of Finance*, 45(4):1089–1108, 1990.

[18] Eugene F. Fama and Kenneth R. French. Dividend yield and expected stock returns. *Journal of Financial Economics*, 22(1):3–25, 1988.

[19] Eugene F. Fama and Kenneth R. French. The cross-section of expected stock returns. *Journal of Finance*, 47(2):427–465, 1992.

[20] Kenneth R. French. Stock returns and the weekend effect. *Journal of Financial Economics*, 8(1):55–69, 1980.

[21] Amit Goyal and Ivo Welch. Predicting the equity premium with dividend ratios. *Management Science*, 49(5):639–654, 2003.

[22] Puneet Handa and Ashish Tiwari. 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, 2006.

[23] Charles J. Jacklin, Allan W. Kleidon, and Paul Pfleiderer. Underestimation of portfolio insurance and the crash of october 1987. *Review of Financial Studies*, 5(1):35–63, 1992.

[24] Narasimhan Jegadeesh. Evidence of predictable behavior of security returns. *Journal of Finance*, 45(3):881–898, 1990.

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

[26] Andrew W. Lo and Craig MacKinlay. Data-snooping biases in tests of financial asset pricing models. *Review of Financial Studies*, 3(3):431–467, 1990.

[27] Narayan Y. Naik, Tarun Ramadorai, and Maria Stromqvist. Capacity constraints and hedge fund strategy returns. *European Financial Management*, 13(2):239–256, 2007.

[28] Hashem Pesaran and Allan Timmermann. Predictability of stock returns: Robustness and economic significance. *Journal of Finance*, 50(4):1201–1228, 1995.

[29] Michael S. Rozeff. Dividend yields are equity risk premiums. *Journal of Porffolio Management*, Fall:68–75, 1984.

[30] Michael S. Rozeff and William R. Kinney. Capital market seasonality: The case of stock returns. *Journal of Financial Economics*, 3(4):379–402, 1976.

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

[32] G. William Schwert. 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), 2003.

[33] Timothy T. Simin. The (poor) predictiveperformance of asset pricing models. *Journal of Financial and Quantitative Analysis*, Forthcoming.

[34] Erik R. Sirri and Peter Tufano. Costly search and mutual fund flows. *Journal of Finance*, 53(5):1589–1622, 1998.

[35] Bruno Solnik. The performance of international asset allocation strategies using conditioning information. *Journal of Empirical Finance*, 1:33–55, 1993.

[36] Dennis Stattman. Book values and stock returns. *The Chicago MBA: A Journal of Selected Papers*, 4:25–45, 1980.

[37] Halbert White. A reality check for data snooping. *Econometrica*, 68(5):1097–1126, 2000.

[38] Yingzi Zhu and Guofu Zhou. Technical analysis and theory of finance. *Working paper*, 2007.