An empirical analysis of initial public offering (IPO) price performance in the United States

Zachary A. Smith
Saint Leo University

Abstract:

For decades, researchers have disagreed about the magnitude and predictability of abnormal securities' price performance generated by initial public offerings (IPOs). The purpose of this study was to identify the best specified and most powerful method of abnormal performance detection and to apply this method to examine the price performance of IPOs. Matched by size, industry, and book-to-market ratios this study explored which of the resulting seven portfolio and matched-firm methods of abnormal performance detection produced the best specified and most powerful test statistics. The paper additionally analyzes IPO price performance to determine if IPOs generate abnormal performance. The researcher used the event study approach for the research design along with the buy and hold abnormal return (BHAR) method of calculating abnormal returns to conduct this analysis. The findings were that (a) all of the matched-firm methods of abnormal performance detection were well specified and powerful (matching by industry affiliation generated the best power and specification result) and (b) that the IPOs generated statistically significant abnormal price performances occurring in: (1) short-term analyses, (2) longer-term analyses, and (3) analyses of the lockup and quiet periods.

Key words: Event study, IPO performance, Quiet period, Lockup period, Specification and power analysis, Short- and long-term abnormal performance, Initial public offering

INTRODUCTION

This research project will provide the reader with a thorough understanding of the anomalies related to IPO price performance, by canvassing the population of IPOs that went public on U.S. financial exchanges from 1985-2008. There were four tests of abnormal IPO performance carried out in this research project, they are tests for: (a) abnormally positive pre-market and initial day of trade performance, (b) abnormally negative longer-term performance, (c) abnormally negative performance occurring around the expiration of the lockup period, and (d) abnormally positive performance occurring during the expiration of the quiet period. In addition to the preceding tests, this study seeks to determine which of seven portfolio-matching (PM) and matched firm (MF) strategies are the best-specified and most powerful estimators of normal performance. The matching strategies evaluated in this analysis were PM techniques by market capitalization, industry affiliation, and market capitalization and book-to-market ratios and MF techniques by market capitalization, industry affiliation, industry affiliation and market capitalization, and market capitalization and book-to-market ratios. This study relied on the use of the event study methodology throughout the analysis and the calculation of abnormal returns by the buy and hold abnormal return (BHAR) method.

The main results and conclusions reached in this analysis were as follows. First, this paper illustrates how poor, in regards to specification and power, PM techniques performed in detecting abnormal performance; on a positive note, all of the MF strategies used to estimate abnormal performance performed remarkably well. Interestingly enough, the MF approach by industry affiliation outperformed the more popular approach—matching by market capitalization and book-to-market ratios. Second, it is apparent that the initial abnormal performance, between the offer and initial trading of shares, is substantial—the current study estimates this abnormal performance at 11.74%—however, abnormal performance is not constrained to pre-market trading. During the initial trading day, IPOs in this sample generated abnormally positive performance of 3.44%. Third, this analysis illustrates that IPOs experience substantial long-term underperformance up to three years after their initial unseasoned equity offering, when compared against firms matched based upon industry affiliation. Finally, IPOs experience significant abnormally positive performance in the five-day period surrounding the expiration of the quiet period of 1.64% and a significantly negative abnormal performance of 1.00% around the expiration of the lockup period.

This paper continues as follows. Section I introduces the theory, empirical work, and conceptual framework of the hypotheses related to IPO performance. Section II presents the proposed methodology. Section III presents the results of the current analysis. Section IV provides a summary of the work and concludes.

LITERARY REVIEW

Studies of IPO performance have been concentrated in two general veins of inquiry: (a) why do IPOs generate abnormal performance and (b) to what extent is this performance abnormal. This project focus on addressing the second of the two preceding questions, namely the significance of this abnormal performance. Many researchers have attempted to answer this question (e.g. Affleck-Graves, Hedge, & Miller, 1996; Ibbotson, 1975; Loughran & Ritter, 2004; Reilly & Hatfield, 1969), but questions regarding their methods used to identify abnormal performance have arisen (e.g. Brav, Geczy, & Gompers, 2000; Brown & Weinstein, 1985; Cheng, Chueng, & Po, 2004; Schultz, 2003). The question that this project seeks to illuminate is

as follows: If IPOs generate abnormal performance, when is this abnormal IPO performance significant and how should academics measure this performance?

METHOD

Traditionally, researchers attempted to measure abnormal performance in research conducted on financial data sets using the event study methodology. This method seems to have been pioneered by Ball and Brown (1968) and Fama, Fisher, Jensen, and Roll (1969), in which the researchers analyze the impact of information on the performance of a publicly traded security; however, according to Campbell, Lo, and MacKinlay (1997) the first published event study was conducted 1933. When the researcher refers to traditional event study designs, like those mentioned in Campbell et al. (1997), normally the researcher has the luxury of (a) an estimation window, (b) event window, and (c) post-event window. However, when dealing with IPOs, the data lacks this estimation period, which is suppose to provide the researcher with normalized expectations of return behavior. Therefore, researchers have to find other means of estimating normal return behavior.

Researchers have used various methods to accomplish this task. In recent research, these methods normally fall into one of two categories: matched firm (MF) approaches or portfolio matching strategies (PM). PM matching strategies have been carried out by Brav and Gompers (1997), Carter, Dark, and Singh (1998), Gompers and Lerner (2005); contrarily, Bhabra and Pettway (2003) and Perfect and Peterson (1997) focused solely on the MF approach; Finally, Ritter, J. (1991) used both MF and PM strategies. Attempts have been made by Barber and Lyon (1997), Mitchell and Stafford (2000), and others to generate tests of different methods analyzing the ability of these different methods to detect abnormal performance—the two general approaches are categorized as event and calendar time analyses.

There are good arguments for and against using various methods; researchers typically prefer one method to the other. This project is interested in evaluating the investor's buy and hold investment experience; therefore, because the buy-and-hold investor's returns are most appropriately modeled in event time, the method used to analyze the investment performance in the current analysis was the buy and hold abnormal return (BHAR) method. The question that is inherently difficult to answer is that if the researcher is interested in analyzing the buy and hold investor's experience, in terms of abnormal performance, what would the best-suited method look like? Mitchell and Stafford (2000) found serious flaws in using the BHAR method to conduct their analysis, but they focused their critiques on PM techniques, using bootstrapping procedures or bootstrapped *t*-statistics, and stated that "BHARS have poor statistical properties, producing biased test statistics in random samples" (p.302). Portfolio-matching techniques produce misspecified test statistics in random samples; however, the method ignored in the Mitchell and Stafford (2000) analysis is the MF approach to benchmarking, which produced well-specified test statistics in the present analysis. Barber and Lyon (1997) concluded that the MF approach, relying on a firm's market capitalization and book-to-market ratios to evaluate firm performance, generated well-specified test statistics throughout their analysis. This project focuses on measuring the buy-and-hold investor's investment experience; therefore, the researcher applied the BHAR method throughout this analysis.

Research Questions

This paper is partitioned into four different research questions; this section develops these questions. The first question that this research project endeavors to answer is which general method, matched-firm or portfolio-matching technique, paired with firm-specific information (i.e. market capitalization, industry affiliation, and book-to-market ratios) provides the best proxy for expected return. The next topic that this project seeks to address is whether short-term abnormal performance occurs, in the process of issuing unseasoned equity shares; the researcher will analyze the specific time horizons in segments, studied by pre-trade and initial trading day results. Lengthening this analysis the project will then seek to determine whether IPOs underperform the market in longer-term analyses. Finally, the researcher will evaluate whether IPOs generate significant abnormal performance in the five-day period surrounding the expiration of the quiet and lockup periods.

Specification and power analyses.

The first portion of the hypothesis testing section will evaluate the performance of potential methods used to identify abnormal performance in similar studies. There has been significant debate regarding whether researchers should use the CAR or BHAR method of calculating abnormal returns when conducting event studies. In the previous subsection, this debate was articulated. In this section, the discussion centers around which method of estimating expected return should be use to conduct event studies, given that the BHAR method is the appropriate method to use to estimate the extent of abnormal performance.

The majority of research projects that attempt to determine which method of abnormal performance detection, PM or MF, to use when conducting event studies conclude that the MF approach works quite well (Ang and Zhang, 2004; Barber and Lyon, 1997). However, researchers seem to continuously revert back to attempting to identify a method of abnormal performance detection that relies on the construction of portfolio benchmarks. In Lyon, Barber, and Tsai (1999) the researchers used skewness adjusted *t* statistics and empirically generated distributions of mean long-term stock returns generated from pseudo portfolios, to compensate for the following biases: (a) the new listing bias, (b) the rebalancing bias, (c) skewness bias, (d) cross-sectional dependence, and/or (e) a bad model problem (p. 197). However, after these efforts are undertaken, the MF approach used to detect abnormal performance, using market capitalization and book-to-market ratio data to match, generated better-specified test statistic than either adjusted portfolio technique.

This analysis will illustrate that the entire set of MF approaches, used to detect abnormal performance, generated well-specified and relatively powerful test statistics—prior to making additional changes to the models of abnormal performance detection. This study assumes that the sample sizes used to conduct the analyses and the 18+ year period that the study was run will minimize the impact of this bias on the results obtained. Finally, this research project provides evidence that when researchers use the MF approach to detect abnormal performance, combined with an independent sampling technique, the model performs very well regardless of the techniques researchers use to match the event firms. In summary, first, there are many biases that researchers can fall prey to when attempting to conduct event studies, second, when researchers use PM techniques instead of MF approaches these biases are magnified, and, third, the biases

affect the results of the PM techniques more than the MF approaches used to detect abnormal performance.

Short-term abnormally positive performance.

The most visible abnormality that currently exists in studies of IPO performance is that IPOs tend to produce extremely abnormally positive performance results a short duration after going public. This excess abnormal return occurs either in the pre-issuance period or in the one-day performance of the post-offering period (see Krigman, Shaw, & Womack, 1999; Loughran & Ritter, 2004; McDonald & Fisher, 1972; Reily & Hatfield, 1969). Miller and Reilly (1987) found that the extent of this underperformance was approximately 9.87% (p. 34) and Ibbotson, Sindelar, and Ritter (1994) reiterated this sentiment by concluding that "first–day returns average 10-15%" (p. 66). Cheng, Cheung, and Po (2004) found, while studying IPO price performance on the Hong Kong financial market, that no trading profits were obtainable once IPOs began trading publicly (p. 853), this finding contrasts those reached in Miller and Reilly (1987), an analysis of IPOs listed in the U.S. markets. Historically, researchers seem to have assumed that IPOs obtained profits in the first trading day. Perhaps, they have ignored the negative social and process implications attached to an empirical finding that the positive IPO performance is constrained to the pre-trading period. If the abnormal performance is constrained between the offer and issuance, then the distributions of shares, and whom the shares are distributed, become a more fundamental question, in regards to affording investors with equal opportunities to profit. This question is relevant because the underwriting syndicate holds an unfair informational advantage over the majority of the investing public.

Long-term underperformance.

Researchers have also provided evidence in support of the theory that IPOs suffer from long-term price underperformance when measured against standard benchmarks (see Affleck-Graves, Hedge, & Miller, 1996; Ibbotson, 1975; Loughran, & Ritter, 1995; Ritter, 1991). Ritter (1989) found that, in his sample of IPOs issued from 1975-84, IPO's 3-year holding period returns (HPR) underperformed portfolios matched based upon market capitalization and industry characteristics by 27.39% (p. 4); Ibbotson, Sindelar, and Ritter (1994) found similar results analyzing IPO data from 1970-1990. Ritter (1989) and Ibbotson (1994) suggested that on average IPOs underperform standard benchmarks from the end of the initial trading day to at least the firm's five-year publicly traded anniversary.

Event-specific Abnormal Performance.

Two events that occur systematically after a company issues unseasoned equity to the public are the expiration of the quiet and lockup periods. Researchers have illustrated that these two events produce abnormal performances in empirical analyses of event studies. However, the directions of the abnormal performances that the two events generate are divergent, and researchers have questioned the magnitude and causes of these abnormal performances. The following two sections will define and review the literature related to the abnormal performance that purportedly occurs during the expiration of the quiet and lockup periods.

At the conclusion of the quiet period, the SEC allows investment firms to initiate coverage of a security. The reason why this period is so interesting is that Bradley, Jordan, and Ritter (2003) have found that from 1996-2000, for all IPOs issued, analysts initiated coverage on 76% of the newly issued IPOs, and of these 76%, analysts initiated coverage on 96% of these issues as a strong buy or a buy (p. 33). This is not what the researcher expected; structurally, the researcher would prefer to see a distribution that, from a probabilistic standpoint, firms rated would just as likely receive a positive rating as a negative rating. According to Bradley et al. (2003), when analysts initiate coverage immediately after the quiet period, the IPOs affected by this event experienced a significantly positive abnormal return of 4.1% in a five-day window surrounding the quiet period (p. 33). If analysts left the newly issued IPOs uncovered at the conclusion of their quiet period, firms experienced an insignificant abnormal return of 0.1% (see Bradley et al., 2003, p. 33). In 2004, Bradley, Jordan, Ritter, and Wolf (2004) attempted to expand this study to include IPOs that went public from January 2001 through July 2002; the impact of the expiration of the quiet period during this time horizon was insignificant (p. 11). In this study, the researcher endeavored to answer why the two research projects differed in regards to their results and analyze whether abnormal performance is significant during the expiration of the quiet period.

Researchers, in the past, have not built a solid case to declare that abnormal performance occurs as the lockup period expires. However, Field, and Hanka (2001) found that from 1988 to 1997, during the expiration of the lockup period, investors experienced a three-day abnormally negative performance of 1.5% (p. 471). The results from Garfinkle, Malkiel, and Bontas (2002) were in agreement with Field et al. (2001), although the Garfinkle et al. (2002) found that negative performance experienced during the expiration of the lockup period was to 4.47%. The two different percentages vary remarkably and the methods that the researchers used to calculate abnormal returns are quite different. A goal of this research project is to add clarity and specificity to this potential anomaly.

METHODOLOGY

A discussion of the rationale behind the decisions to use the BHAR method to calculate abnormal returns over the choice of the CAR method was address earlier in this document; this section will describe how the researcher will implement the method and run the power and specification analyses. Furthermore, the sample sizes are different in many of the analyses, even though the study canvassed the entire time horizon, from January 1985 to December 2008; because of incomplete records, the researcher conducted some of the analyses with much smaller sample sizes than originally anticipated. This section will review the methodological procedures applied to conduct the power and specification analyses as well as the test of abnormal performance and the samples sizes of each test.

Power and Specification Analyses

In this analysis, the researcher conducted the power and specification analyses based partially on the methodology described in Ang and Zhang (2004). The following paragraphs describe the adjustments made to the Ang et al. (2004) methodology. It is appropriate to note here that two different power and specification analyses were run to determine how increases in sample sizes would influence the metrics ability two identify abnormal performance. In the first

analysis, the researcher took 10 non-repeating samples of 50 companies taken from the list of Russell 3000 constituents each year of this analysis. Next, these 10 yearly samples of 50 companies were combined to produce larger samples (n = 500).

To evaluate which of the seven different benchmarking techniques generated the best specified test statistics, the project needs a pool of random companies to evaluate the metric used as a proxy for normal performance, a method used to calculate abnormal performance, and a method to aggregate abnormal performances across the sample. This analysis uses the BHAR methodology to calculate abnormal performance and seven different methods based upon either portfolio matching or matched-firm methods used to proxy for expected returns.

This project used two different procedures to obtain proxies for expected returns: (a) PM and (b) MF. For the MF approach, the researcher extracted the sample firms used in this analysis from a list of the components of the Russell 3000 Index each year. If a company was included in the Russell 3000 list of companies, the company was eligible to be a MF in this analysis—each year, the list was updated, from 1985-2002, due to the addition and deletion of firms from the list of constituents each year. If a firm is matched based upon any singular firm characteristic (i.e. market capitalization or industry affiliation), a pool of potential matched firms are identified based upon this criteria, then these firms are then sorted and the closet match is selected; furthermore, if there are multiple firms that meet the matching requirements, a number is assigned to each potential match and a firm is randomly selected from the potential matches. If two factors are included in the matching procedure (i.e. industry and market capitalization and market capitalization and book-to-market ratios), the firms are sorted by the most appropriate factor (i.e. industry affiliation for the industry/market capitalization sort and market capitalization for the market capitalization and book-to-market sort) first, and then they are sorted based upon the second factor.

When this research project used the PM technique, it relied on external portfolios to match the firm to a portfolio with a similar likeness. The procedure for matching was simple: the researcher paired the event firm with a portfolio grouping compiled and maintained on Dr. Kenneth French's website (URL: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). The researcher then carried out this matching procedure for PM approaches based upon industry affiliation, market capitalization, and market capitalization and book-to-market ratios.

After the pairings were made, the performance of the simulated event firm (randomly drawn from the Russell 3000 list each year) and the matched firm or portfolio match were compared, in terms of specification, for 1, 2, 3, and 4-year time horizons. The researcher then ran the specification analysis to determine if a metric would conclude that abnormal performance had occurred when in actuality it had not—throughout this analysis the level of significance was set at 5%. Therefore, for each pairing the following Buy and Hold Abnormal Return (BHAR) was calculated:

$$BHAR_{i,t} = \prod_{t=1}^{\tau}\left[1 + R_{i,t}\right] - \prod_{t-1}^{\tau}[1 + E\left(R_{i,t}\right)] \quad (1)$$

The researcher then pooled the result of this formula for each sample taken for this analysis and then compiled the following summary statistics: (a) sample size, (b) sample average, (c) sample standard deviation.

After the researcher compiled these statistics for each sample taken, he calculated the following statistic, Barber and Lyon (1997):

$$t_{BHAR} = \frac{\overline{BHAR_{i,t}}}{\sigma(BHAR_{i,t})/\sqrt{n}} \quad (2)$$

The researcher then took the *t* statistics and grouped them based upon their respective sample to generate the empirical size (ES) statistic. The researcher calculated this statistic by taking each sample of 50 or 500 observations contained in the yearly cohorts, summing the number of times that the given metric identified abnormal performance, and dividing this sum by total number of observations contained in the cohort.

The power analysis uses the results of the specification analysis as a base to continue the evaluation of the given method of abnormal performance detection. All models of normal performance that have properly constructed should identify no abnormal performance, given that enough random draws were taken from a randomly selected population of return data. From a base of zero abnormal performance, abnormal performance is simulated across the entire sample by taking the average performance add adding either positive or negative percentage movements of 1, 5, 10, 15, 20, 30, 50, and 75%.

Using the outcome from these simulations, this project then calculates the t_{BHAR} for each level of simulated abnormal performance. The researcher expects abnormal performance to be negligible where zero abnormal performance is simulated and increase as both positive and negative abnormal performances are simulated; therefore, when charting the results of this analysis the researcher is looking to obtain a v or u-shaped power curve, centered on zero abnormal performance and increasing substantially as abnormal performance is simulated. After simulating abnormal performance, for each simulation and each metric, the Empirical Power (EP) statistic was calculated. The EP statistic is similar to the ES statistic, but when the EP statistic is calculated the researcher is analyzing how well a metric identifies the researcher's simulation of abnormal performance and when it does not. Thinking back to the u or v-shaped curve, at zero percent simulated abnormal performance the researcher wants to see that the metric does not identify abnormal performance, therefore, the ES statistic would be zero, but at each increment (positive and negative) from 1% to ∞, the researcher wants to see that the metric's ability to detect abnormal performance increases.

Short-term abnormal performance.

The time horizon used to evaluate short-term abnormal performance shrank significantly when compared to the other studies in this analysis. Using the sources that were available (Hoovers IPO Central, Edgar IPO), this study was able to obtain premarket offering prices for the IPOs included in this analysis. The researcher used the following time horizon, January 1, 1997 to December 22, 2005, for tests conducted on the performance of the initial day of public trading and April 12, 1996 to January 28, 2008 for tests conducted on pre-trade performance. Although, this is a substantial reduction in the intended sample, there were still a significant number of observations in each sample—the researcher identified 1,876 observations for premarket performance and 2,143 observations for the initial day of trading. Even if it was possible to obtain performance data prior to January 1, 1997 for the initial day of trading in IPOs, the CRSP database, which was used to obtain daily pricing data in this analysis, did not have initial day trading data for IPOs listed prior to January 1997.

It is important to note, in the analysis of pre-trade performance, the public does not openly share their return expectations and the research project lacks a specific time horizons (e.g.

does the offering to issue period last 12 hours, 24 hours, 36 hours, or more) to match the return of the event firm against. Therefore, the researcher compared the aggregate returns obtained in the pre-public trading period with the returns obtained by investor's investing in a market proxy—the researcher uses standard market indices to obtain this performance (e.g. Russell 3000, S & P 500, NASDAQ Index, etc.). Therefore, when the researcher evaluated abnormal performances occurring in pre-public trading he started with the assumption that the aggregate IPO will produce a return of 0%, and compared this return against the return of the DJIA, Russell 3000, and NASDAQ indices, to gain some insight on how substantial premarket IPO performance is. The researcher will revert to using the best-specified and most powerful method used to detect abnormal performance in the remainder of the analyses, because the researcher has obtained public trading data that can be compared to the event firm's performance. Therefore, in the initial day of trading, the researcher will take the returns obtained in from the sample of IPOs and, using the BHAR method to detect abnormal performance, match these firm's to the best-specified and most powerful metric identified in the preceding power and simulation analysis to determine whether abnormal performance has occurred.

Long-term abnormal performance.

To obtain a general sample to run tests for longer-term abnormal IPO performance, the researcher used the Field-Ritter dataset of founding dates, identified in Loughran and Ritter (2004; as noted in http://bear.cba.ufl.edu/ritter/foundingdates.htm) for companies that went public from 1985 to 1996. Additionally, the researcher obtained information pertaining to IPO issuance from 1996 to 2002 from on-line IPO databases (e.g. Hoovers IPO Central, Edgar IPO). The total sample of IPOs used in this analysis was 5,883. Any company that had an offer price of less than $5 or were foreign offerings were removed from this list—this is the base IPO list used for both the analysis of longer-term abnormal IPO performance and event specific abnormal IPO performance. Using the best-specified and most power method of abnormal performance detection, the researcher paired the IPOs in this sample with the benchmark to determine whether abnormal performance is evident using time horizons ranging from day 2 to trading day 750.

Event-Specific Abnormal IPO Performance

For tests of abnormal performance occurring during the expiration of the quiet and lockup periods, the sample size shrank to 5,529 due to firm attrition. In this analysis, the event horizon was the five-day period surrounding the day of the specific event—either the expiration of the lockup period or the conclusion of the quiet period. The researcher compared the BHAR obtained from the IPO experiencing the event against the benchmark and the results of these individual analyses were aggregated to give an average BHAR for the entire sample of IPOs issuing shares over this period.

RESULTS

This section provides the results of the tests that the researcher conducted to indentify abnormal price performance related to the issuance of unseasoned IPO issuance. The first section provides the results of the specification and power tests the researcher conducted on seven metrics used to identify abnormal performance. Sections 2 through 4 will display the results of

tests that the researcher conducted to identify abnormal performance, using the best-specified and most power testing procedure.

Specification and Power

The purpose of this section was to determine, which method of benchmarking was most effective in testing for abnormal performance during the sample time horizon. Based upon the review of literature, the researcher employed two broad methodological strategies to conduct the specification and power analyses--the portfolio matching and the matched-firm approaches. The first subsection will present the specification results and the second subsection will present the results of the power analysis.

Specification Analysis.

The first question that this analysis answered is as follows: in samples of 50 and 500 companies, how often did the randomly drawn event firm (i.e. drawn from the list of Russell 3000 constituents each year) generate statistically significant abnormal performance. After the researcher conducted the hypothesis test for each sample, the number of rejections were added together and divided by the number of observations, thus resulting in the *ES* statistics. The researcher displayed the results of the specification analysis Table 1.

The researcher found that all of the approaches using the MF technique (i.e. matching based upon market capitalization, industry affiliation, industry affiliation and market capitalization, market capitalization and book-to-market ratios) were generally well specified, using a level of significance of 5%. The MF approach based upon market capitalization, alone, did incorrectly identify abnormal performance in 5.56% of its samples, using sample sizes of 50 and 11.11% with sample size of 500.

To determine if the metrics were misspecified, the researcher conducted an additional test to determine whether the *ES* was significantly different from the theoretical 5% level of significance—where α was the level of significance and n was the sample size. The *ES* interval ranges from 1.82% to 8.18% for the 180 samples of 50 companies and from negative 5.07% to 15.07% in the 18 samples of 500 companies; therefore, the percentage of the observations in the samples that were previously rejected were within our error boundaries. MF approaches, matched based upon Market Capitalization, Industry Affiliation and Market Capitalization, and Market Capitalization and Book-to-Market Ratios Companies, generated spurious rejections; however, these rejections where not statistically different than the theoretical level of significance. Even though they were not statistically different from the theoretical level of significance used in this analysis, they were different. The best-specified MF approach used in this analysis and the approach that did not identify abnormal performance greater than the theoretical level of significance in any of the analyses was the MF technique based solely upon industry affiliation.

Each of the PM strategies (i.e. matched by market capitalization, industry affiliation, and market capitalization and book-to-market ratios) rejected the null hypothesis; this indicates an identification of abnormal performance even though the researcher had not simulated abnormal performance. Every specification test, using the PM techniques, regardless of how it was matched to the event firm, generated misspecified test statistics and in all cases was significantly different from the theoretical level of significance.

As illustrated in Table 1, as the sample size increases from 50 to 500, the observed percentage of spurious rejections decreased using the MF approach. This occurs because, without simulating abnormal performance, researchers would expect to detect no abnormal performance. Given the preceding results, the researcher found and believes that it is evident that the MF approach is a better-specified method of abnormal performance detection than the PM strategies. The researcher concluded that the best-specified MF approach is implemented using MF strategy, matching by industry affiliation.

Power Analysis.

The purpose of power analysis was to determine which method had the least type II error, and which methodology had the highest power. This research project relied on running the power analysis by simulating abnormal performance in +/- .01, .05, .10, .15, .20, .30, .50, and .75 intervals to the individual BHARs derived from the results of the specification analysis. In essence, this analysis forced the average abnormal performance away from zero and imposed abnormal performance on the BHAR. The researcher calculated the *EP* statistic by adding each of the sample average BHARs, for each level of simulated abnormal performance, and dividing this sample average by the size of each sample. Again, the researcher obtained 180 samples of 50 observations in the first round of the analysis and 18 samples of 500 companies in the second round of the analysis.

All of the MF approaches had defined power curves—the traditional U or V shaped—the power curve are centered approximately centered on zero, the point where no abnormal performance is simulated. In comparison, the PM benchmarks had no defined structure or at least not the structure needed to make credible inferences pertaining to the power of the benchmark. Again, the PM benchmarks failed to approach acceptable standards that are necessary to judge the benchmarks ability to detect abnormal performance; in the remaining analyses, the PM techniques were not included because the researcher did not considered them to be meaningful alternatives to the MF approach.

If the researcher simulated abnormal returns of 15%, the competing matched-firm approaches only rejected the null hypothesis (identifying abnormal performance) in approximately 30% of samples using sample sizes of 50 observations. When the sample is expanded to 500 from 50, the MF approach identified abnormal performance in 80% of the samples. Therefore, as the sample size increases, the power curve narrows making the employed methodology appropriate.

There is still no statistically significant difference between the various MF approaches to benchmarking. When conducting the remainder of the tests the research project was concerned with the speed at which the metric deteriorates. As the event horizon increased the method's ability to detect abnormal performance decreases. Comparing the event horizons of one-, two-, three-, and four-years using sample sizes of 50, this study found that a simulated abnormal performance of +/- 10% will be detected in 17%, 10%, 6%, and 6% of the samples; in samples of 500 observations the percentage of detection are 80%, 55%, and 25%, and 11%, respectively. To analyze the general ability of each of the MF approaches to detect abnormal performance, this project now will identify when the metrics identify abnormal performance in 95% of the analyses. The *EP* reached 95% at 15%, 15%, 30%, and ~40% of simulated abnormal performance using an event horizon of one-, two-, three-, and four-years, respectively, and sample sizes of 500 observations. Therefore, if researchers intend on using the matched firm

approaches identified in this analysis their sample sizes and predicted level of abnormal performance should be significantly large.

Initial Performance

The following section focuses on detecting abnormal performance during the initial trading period. The main questions posited in the following section was whether unseasoned IPOs produced abnormal performances in the time preceding public trading and if this abnormal performance continued into the first day of public trading. The results of the analysis conducted prior to public trading are reported first and then an analysis of whether IPOs produced abnormal performances on their first day of trading is reported.

IPO Performance (Pre-issuance)

This project uses the average returns in this round of the analysis; there is no way to pair event firms with another firm base upon firm specific criteria, because this performance occurs prior to public trading. The average return that IPOs generated prior to public trading or from their offering to their issuance to the public was 11.74%, with a sample standard deviation of 31.16%, and 1876 observations taken from April 12, 1996 to January 29, 2008. The researcher conducted a *t* test to determine if the 11.74% performance was statistically different from zero. The resulting *t* statistic was 16.32, which was outside the critical value of 1.645 for a one-tailed statistical test, given a 5% level of significance.

The preceding analysis illustrated the difference between the performance obtained by IPOs pre-public trading and an expectation of zero abnormal performance. Since this is the pre-public trading period, there is no specific way to pair the individual IPO performance with a benchmark. Therefore, the researcher aggregated the returns into monthly IPO cohorts, these performance cohorts assumes that the investor obtains shares of the IPO in the offering and sells the shares at the initial trade on the first day of public trading. In Table 2, the researcher has illustrated how abnormal IPOs performance is in pre-public trading.

To make this analysis comparable to the results obtained in the remainder of the analyses contained in this project the researcher paired these returns with the performances of standard benchmarks over our time horizon. The researcher displayed the results of these comparisons in Table 3. Table 3 shows the average monthly performance of IPO cohort versus those of DJIA, Russell 3000, and the NASDAQ Composite Indices over the period analyzed. As the numbers in Table 3 indicate, at 5% level of significance for a one-tail *t* test (*t* critical of 1.66), the researcher rejected the null hypothesis for only the IPO sample, implying that the IPO group experienced significant abnormally positive returns. None of the benchmark indices produced abnormal returns.

The DJIA was the best performing benchmark out of the three potential benchmarks chosen for this analysis; the project continues to analyze whether the IPO cohort significantly outperformed the best performing index, which was the DJIA in this period. The average difference between the IPO cohort and the DJIA's yearly average return was 8.41%, with a sample standard deviation of 13.86%, and observations' occurring over 139 months--the computed *t* statistics was 7.15. Again, with a 95% level of significance for a one-tailed test the critical value of *t* is 1.66; therefore, this research project rejects the null hypothesis and identifies

statistically significant evidence that abnormal performance occurred during the pre-public trading period when compared against standard indices.

Initial Day of Public Trading

The next analysis determines whether IPOs generate abnormal performance on the first day of public trading. To answer the question, the analysis evaluates the returns of IPOs issued to the public from January 1, 1997 to December 22, 2005, the sample contains 2,143 observations. Using a standard *t* test, this analysis uncovered the following: the average return across the IPOs was 3.44% and the average performance of the matched-firm benchmark was 0.13%. The sample standard deviation was 16.27%; resulting in a *t* value of 9.423, which when compared to a critical value of 1.645, at a 95% level of significance, indicated that the IPOs abnormal returns on the first day of trade are statistically significant. The returns of IPOs on the first day of trade are significantly different from the returns obtained for the MF benchmark.

Long-term Abnormal Performance

This round of the analysis turns to evaluating whether significant abnormal performances occur after the short-term abnormal performances. This project accomplished its longer-term analysis by canvassing the population of IPOs issued in the U.S. from January 1, 1985 to December 31, 2002. The study identified 5,583 IPOs to use in this analysis; the researcher matched these IPO based upon industry affiliation to a benchmark firm. The BHAR was calculated and the researcher identified the sample average and standard deviation given the individual BHARs. The output, which encompasses trading day 2 through 750, is the averaged BHAR across the entire sample over the specified time horizon. The researcher evaluated the data and generated a two-tailed *t* test for all 749 time-horizons.

The analysis of the data illustrates that, from trading days 5 through 12 IPOs significantly underperformed the MF benchmark, at day 17 the trend changed positive, and it was statistically significantly positive until trading day number 120 (with one insignificant reading on day 33)--at day 120 the BHAR was 1.934%. The averaged BHAR continued along insignificantly, but positive, until reaching trading day 161. However, the BHAR did not generate a significantly negative BHAR until it reached 201 trading day. The BHAR remained significantly negative through the remainder of the analysis. Moreover, at the end of year three the highest abnormally negative performance occurred, which was -22.41%.

Quiet and Lockup Expiration

To construct a test for abnormal performance at the expiration of the lockup and quiet periods this project canvasses the same population of IPOs used in the longer-term analysis. The number of observations for the quiet and lockup period analyses was 5529. To carry out these analyses this section calculates the 5-day BHAR surrounding the date in which the quiet period ended and the lockup period expired.

Quiet Period.

For the analysis of performance surrounding the expiration of the quiet period, the sample average BHAR was 1.64%, for the five-day period surrounding the event and the sample standard deviation was 13.9%. The resulting *t* statistic was 8.75, using a 95% level of significance the critical value was 1.645; the null hypothesis is rejected—at the conclusion of the quiet period IPOs produce a significantly positive abnormal performance.

Lockup Expiration.

In the analysis of the performance resulting from the expiration of the lockup period, the researcher found significantly negative performance of 1.00%. In addition, the sample standard deviation was 13.74%, therefore, the resulting *t* test produced a test statistic of –5.41, and with a 5% level of significance the critical *t* value is -1.645. Therefore, again, the researcher rejected the null hypothesis and concluded that significant negative abnormal performance of 1.00% occurred at the expiration of the lockup period.

SUMMARY & CONCLUSION

In the preceding section, the researchers has (a) presented a well specified and powerful method used to identify abnormal performance when conducting event studies, (b) shown that short-term abnormal IPO performance is positive, (c) illustrated that events occurring throughout the IPO process instigate abnormal performances, and (d) provided a description of IPO performance over the initial three years of seasoning. The results of the analyses related to event specific performances--abnormal performances occurring at the expiration of the quiet and lockup periods--generated significant, but not substantial abnormal performance. However, the pre-public trade abnormal performance of 11% and 3% abnormal performance in the initial trading day, together with long-term underperformance of IPOs in excess of 30%, seem to suggest that substantial performance abnormalities occur when companies issue unseasoned equity shares to the public.

Researchers focus the majority of their explanations that attempt to explain why short-term abnormal performance occurs on the asymmetric information hypothesis. To summarize, according to Ritter and Welch (2002), either investors are more informed than the issuer about the market demand for the company's shares or the investor believes that the issuer knows more about the firm's prospects and need protection against potential market lemons (IPOs that underperform). Purnanandam and Swaminathan (2004) questioned the conventional wisdom that companies initially discount their shares when they offer them to the public, for whatever reason. Purnanandam et al. (2004) found that, in a sample of over 2,000 IPOs issued from 1980 to 1997, companies typically overpriced IPOs, when the researchers compared these IPOs to their non-IPO counterparts the over pricing ranged from 15% to 50%, depending on the matching criteria. Puranandam et al. provided the first real critique of what has become general knowledge in the academic community: Companies typically under price their shares when they issue unseasoned equity. If IPOs are initially overpriced and this overpricing increases—not only in the period prior to public trading, but IPOs continue to generate significantly positive abnormal performance in their first day of trading—does this signal market inefficiency?

It would be a mistake to conclude that empirical evidence supports the conjecture that markets are inefficient. However, this initial over-pricing, followed by substantial short-term abnormally positive performance, which is followed by—over a period of three years—a reversal to longer-term underperformance could at least hint at market inefficiency. Efficient market theory concedes that short-term departures from fundamental or intrinsic will exist in the marketplace; however, prices will rapidly adjust and the market will eliminate pricing discrepancies. In the longer term analysis of IPO performance, the researcher found that when IPOs are trading under their lockup provision, the returns are generally positive. However, as the IPOs approach the expiration of the lockup period the performances generated by the IPOs evaluated in this analysis were resoundingly negative.

The expiration of the lockup period occurs at approximately trading day number 128 (i.e. 180 calendar day lockup period is equivalent to ~26 weeks, subtracting the weekends equals 128 trading days). In this research project's ex post analysis, after testing all ex ante hypotheses, it became apparent that the downward trend in IPO prices, following the expiration of the lockup period, was remarkable. From trading day 128 to 350, which the researcher has approximated at 241 calendar days—one-year, there was a decline of .05% every trading day when compared against a firm matched based upon industry affiliation. The regression summary is appealing, the *R* Squared value was in excess of .98 and the relationship was very significant ($p = .001$). The trend is undeniable and significant. After the IPO researches it's lockup expiration, it is likley to experience a downward trend of losing approximately .05 percentage points in value each day that it trades for approximately one-year.

The general conclusion that the researcher has reached in this analysis is as follows, "when it comes to participating in the IPO market, buyer beware." First, and foremost, the process of issuance is not fair, there are not fair opportunities for economic profit. A class of sophisticated investors reap the benefits of the 11.74% of performance occuring prior to public trading and in the initial trading day investors may be able to obtain approximately 3 percentage points of positive performance, however, the investors have to buy at the market open and sell at the closing price on the security's initial trading day. If the average investor does not sell at the market close, holding onto the newly issued security will generate a negative 3% price movement from trading day 2 through trading day 7. This is then followed by a substantial upswing in performance and, ofcourse, eventually if held long enough investors will feel the sting of longer-term negative abnormal performance of 22.41% after approximately three years. The researcher has provided investors an overview of the patterns that IPOs seem to have exhibit from 1985 to 2008; hopefully, the average investor finds a meaningful way to put this infomration to use.

REFERENCES

Affleck-Graves, J., Hegde, S., & Miller, R. E. (1996). Conditional price trends in the aftermarket for initial public offerings. *Financial Management, 25(4)*, 25-40.

Ang, J. S., and Zhang, S. (2004). An evaluation of testing procedures for long horizon event studies. *Review of Quantitative Finance and Accounting, 23*: 251-274.

Ball, R., & Brown, P. (1968). An empirical evaluation of accounting income numbers. *Journal of Accounting Research*, 159-178.

Barber, B., and Lyon, J. (1996). Detecting abnormal operating performance: The empirical power and specification of test statistics. *Journal of Financial Economics, 41*: 359-399.

Barber, B., and Lyon, J. (1997). Detecting long-run abnormal stock returns: The empirical power and specification of test statistics. *Journal of Financial Economics, 43*: 341-372.

Bhabra, H. S., & Pettway, R. H. (2003). IPO prospectus information and subsequent performance. *The Financial Review*, *38,* 369-397.

Brav, A., Geczy, C., and Gompers, P. (2000). Is abnormal performance following equity issuances anomalous? *Journal of Financial Economics, 56(2)*: 209-249.

Brown, S., and Warner, J. (1985). Using daily stock returns: The case of event studies. *Journal of Financial Economics, 14*: 3-31.

Brown, S. J., & Weinstein, M. (1985). Derived factors in event studies. *Journal of Financial Economics*, *14(3),* 491-5.

Campbell, J. Y., Lo, A. W., & MacKinlay, A. C. (1997). *The econometrics of financial markets*. Princeton, NJ: Princeton University Press.

Carter, R. B., Dark, F. H., & Singh, A. K. (1998). Underwriter reputation, initial returns, and the long-run performance of IPO stocks. *The Journal of Finance*, *53(1)*, 285-311.

Cheng, W., Cheung, Y. L., & Po, K. (2004). A note on the intraday patterns of initial public offerings: Evidence from Hong Kong. *Journal of Business Finance & Accounting*, *31 (5-6)*, 837-860.

Fama, E. F, Fisher, L., Jensen, M. C., & Roll, R. (1969). The adjustment of stock prices to new information. *International Economic Review*, *10(1),* 1-21.

Gompers, P. A., & Lerner, J. (2003). The really long-run performance of Initial Public Offerings: The pre-NASDAQ evidence. *The Journal of Finance, 57(4)*, 1355-1392.

Ibbotson, R. G. (1975). Price performance of common stock new issues. *Journal of Financial Economics (2)*, 235-272.

Kothari, S., and Warner, J. (1997). Measuring long-horizon security price performance. *Journal of Financial Economics,* 43: 301-339.

Krigman, L., Shaw, W., and Womack, K. (1999). The persistence of IPO mispricing and the predictive power of flipping. *The Journal of Finance, 54(3)*, 1015-1044.

Loughran, T. and Ritter, J. (2004). Why has IPO underpricing changed over time? *Financial Management, 33(3),* 5-37.

Lyon, J., Barber, B., and Tsai, C. (1999). Improved methods for tests of long-run abnormal stock returns. *The Journal of Finance*, *54(1)*: 165-201.

McDonald, J. and Fisher, A. (1972). New-issue stock price behavior. *Journal of Finance, 27(1),* 97-102.

Perfect, S. B., & Peterson, D. R. (1997). Day-of-the-week effects in the long-run performance of initial public offerings. *The Financial Review, 32(1)*, 49-70.

Purnanandum, A. K., & Swaminatham, B. (2004). Are IPOs really underpriced? *Review of Financial Studies, 17(3)*, 811-848.
Reilly, F. and Hatfield, K. (1969). Investor experience with new stock issues. *Financial Analysts Journal, 25(5),* 73-80.
Ritter, J. R. (1991). The long-run performance of initial public offerings. *The Journal of Finance, 45(1)*, 3-27.
Ritter, J. R., and Welch, I. (2002). A Review of IPO Activity, Pricing, and Allocations. *The Journal of Finance*, *57(4)*, 1795-1828.
Schultz, P. (2003). Pseudo market timing and the long-run underperformance of IPOs. *The Journal of Finance, 58(2)*, 483-517.

TABLES & FIGURES

Table 1. The results of the specification analysis conducted on 180 samples of 50 companies and 18 samples of 500 companies.

Specification Analysis							
180 Samples of 50 Simulated Event Firms							
			Matched Firm			Portfolio Match	
Years	Mcap	Ind	Ind & MCap	MCap & B toM	Mcap	Mcap & B to M	Industry
1	5.00%	3.89%	4.44%	3.33%	43.89%	44.44%	46.67%
2	2.78%	1.67%	3.33%	2.78%	31.67%	25.56%	56.67%
3	2.20%	1.67%	2.22%	1.11%	33.33%	28.89%	65.56%
4	5.56%	3.89%	3.89%	2.78%	47.22%	36.67%	79.44%
18 Samples of 500 Simulated Event Firms							
			Matched Firm			Portfolio Match	
Years	Mcap	Ind	Ind & MCap	MCap & B toM	MCap	MCap & B to M	Industry
1	0.00%	0.00%	0.00%	0.00%	83.33%	83.33%	66.67%
2	0.00%	0.00%	0.00%	5.56%	66.67%	66.67%	83.33%
3	11.11%	0.00%	0.00%	5.56%	66.67%	77.78%	94.44%
4	0.00%	0.00%	5.56%	0.00%	66.67%	61.11%	94.44%

Notes: Mcap - Market Capitalization, Ind - Industry Affiliation, B to M - Book to Market Ratio.

Table 2: Pre-Public Issuance IPO Returns.

Pre-Public Trade - Average Monthly IPO Performance from 1997 to 2007					
Year	Average Return	Standard Deviation of Returns	Number of Months	*t* value	$t_{Critical}$ - 1.796
2007	11.94%	6.34%	12	6.52	Reject
2006	9.72%	5.00%	12	6.73	Reject
2005	9.75%	5.02%	12	6.73	Reject
2004	9.25%	4.12%	12	7.77	Reject
2003	7.14%	6.68%	12	3.70	Reject
2002	5.52%	4.09%	12	4.68	Reject
2001	7.79%	6.30%	12	4.29	Reject
2000	34.60%	32.56%	12	3.68	Reject
1999	5.23%	7.33%	12	2.47	Reject
1998	-0.08%	4.41%	12	-0.06	Accept
1997	1.55%	2.16%	12	2.49	Reject

Notes: Table 2, provides the yearly returns of IPOs assuming that an investor was issued shares of each IPO and subsequently sold those shares on the market when the IPO began trading publicly.

Table 3: Average Monthly IPO Performance Compared Against Standard Benchmarks

Sample Average Return Comparison from July 1996 to January 2008 Test Ho </= 0, H1 > 0, $t_{Critical}$=~2.61 (*t→.005)*				
	IPO	DJIA	Russell 3000	Nasdaq
Sample Average Return	8.96%	0.55%	0.46%	0.23%
Standard Deviation	13.52%	4.04%	4.42%	8.19%
Count	139	139	139	139
T-Value	7.81	1.60	1.22	0.33
Accept/Reject Decision	Reject	Accept	Accept	Accept

Notes: Table 3 provides an analysis of average yearly returns for a strategy that invests in every IPO that went public from July 1996 to January 2008 and compares the IPO performance result against standard benchmarks