2000

Essays on the Gain -Loss Pricing Model.

Kevin Chun-hsiung Chiang
Louisiana State University and Agricultural & Mechanical College

ESSAYS ON THE GAIN-LOSS PRICING MODEL

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Interdepartmental Program in Business Administration (Finance)

by
Kevin Chun-Hsiung Chiang
Diploma, National Taipei Institute of Technology, 1986
M.B.A., Mankato State University, 1994
August 2000

UMI Number: 9984318

Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

ACKNOWLEDGMENTS

I am deeply indebted to my committee chairman, Dr. Ji-Chai Lin, for his dedication, effort, and guidance throughout my doctoral program and in completing this dissertation. I would also like to thank the members of my dissertation committee, Drs. Faik Koray, Kelly Pace, and Gary Sanger for their valuable insights and comments into this research.

I express my sincerest gratitude to my mentor, Dr. George Frankfurter. I thank Jay Ritter for the IPO dataset used in the third chapter. Special thanks are to my fellow doctoral students and staff members in Finance department. Finally, I would like to dedicate this work to my father, Chin-Chi Chiang, to my mother, Mei-Zen Young, to my wife, Emily Huang, and to my son, Alan Chiang. Without their love, I would not be able to accomplish this work.

TABLE OF CONTENTS

ABSTRACT

This dissertation conducts empirical examinations of a new normative (equilibrium) model, the Gain-Loss Pricing Model (GLPM) of Lin (1999a), in which loss aversion is intuitively incorporated into investors' portfolio decisions. In this equilibrium, the risk-return relation is based on the tradeoff between the expected market-related gain of an asset and its expected market-related loss. In addition to its economic intuition, the new model is shown to be more robust than the mean-variance-based Capital Asset Pricing Model (CAPM).

This dissertation consists of two essays. The first essay examines the empirical power of the GLPM using NYSE/AMEX/NASDAQ stocks. This testing framework has a high testing standard that investigates whether there is a systematic component of asset returns left unexplained by the new model and places no restrictions on the sampling distribution of the statistic. The test results indicate that no more than 3% of sample stocks are mispriced according to the model during any given five-year test period in the 1948-1997 sample period. We also find that the mispricing in small size portfolios is not severe. The evidence implies that most sample firms are priced such that their risk-return relation is consistent with the GLPM.

Based on these testing results, the second essay proposes a long-term performance evaluation framework. This framework is capable of mitigating the skewness problem of long-term abnormal return distributions and avoiding the aggregation problems in many long-term performance tests. The specification of the long-term performance is evaluated using samples of randomly selected

NYSE/AMEX/NASDAQ stocks and simulated random event dates. Simulation results show that the long-term performance evaluation framework based on the GLPM is well-specified.

CHAPTER 1

INTRODUCTION

Based on the Capital Asset Pricing Model (CAPM), researchers have identified a number of anomalies in asset prices.[1] The anomalies are inconsistent with the joint hypotheses that the market is efficient and that the CAPM is a correct model. While it is possible that the market may not be perfectly efficient, many researchers, including Banz (1981), Reinganum (1981), Jegadeesh (1992), and Fama and French (1992, 1993, 1996), argue that the CAPM is misspecified. Furthermore, on theoretical grounds, Dybvig and Ingersoll (1982) and Jarrow and Madan (1997) show that CAPM pricing allows arbitrage opportunities in any market with traded options and suggest that the CAPM has serious theoretical drawbacks.

Recently, Lin (1999a) proposes a new equilibrium asset pricing model, the Gain-Loss Pricing Model (GLPM), which is shown to be more robust than the mean-variance-based CAPM. In addition, the GLPM intuitively incorporates loss aversion into equilibrium asset pricing. It is well known that economic agents are averse to losses when they make risky choice involving monetary outcomes (Kahneman and Tversky (1979), Tversky and Kahneman (1986, 1991, 1992)).

Since the GLPM has economic intuition and robustness, two questions naturally arise. First, can it empirically explain asset returns? Currently, empirical tests based on normative (equilibrium) asset pricing models have not been able to provide convincing results. On the other hand, while positive (empirically based)

[1] The anomalies include the price-earning-ratio effect (Basu, 1977), the January effect (Keim, 1983), the size effect (Banz, 1981; Reinganum, 1981), market overreaction (De Bondt and Thaler, 1985), market underreaction (Jegadeesh and Titman, 1993), and others.

asset pricing models tend to fit return data better, they are *ad hoc* in nature. As Ferson and Harvey (1999) state, "empirical asset pricing is in a state of turmoil." Their statement seems to call for studies that can empirically show that asset returns follow a well-specified theoretical model.

The second question is whether the GLPM is useful in evaluating long-term performance. Recent empirical studies document a number of anomalies in long-term performance. [2] The empirical results bring forth the resurgence of the debate on market efficiency. However, traditional long-term performance studies suffer from theoretical and empirical difficulties that restrain their usefulness as tests of the efficient market hypothesis. Specifically, Loughran and Ritter (2000) argue that tests of market efficiency require a normative (equilibrium) asset pricing model. According to Loughran and Ritter, "if a positive (empirically based) model is used, one is not testing market efficiency; instead, one is merely testing whether any patterns that exist are being captured by other known patterns." An example of positive asset pricing models is the Fama-French (1993) three-factor model. Furthermore, long-term performance evaluation is sensitive to the asset pricing model employed (Fama (1998)). Ideally, to accurately evaluate long-term performance, one needs a well-specified asset pricing model. Therefore, it is important to know whether the long-term performance evaluation framework based

[2] These anomalies include market overreaction (De Bondt and Thaler, 1985), the long-term underperformance of initial public offerings (Ritter, 1991), market underreaction (Jegadeesh and Titman, 1993), and the long-term underperformance of seasoned equity offerings (Loughran and Ritter (1995), Spiess and Affleck-Graves (1995)).

on the normative (equilibrium) GLPM is well-specified. It is, also, of interest to see how long-term anomalies behave under the GLPM.

Since the potential contributions of the GLPM to the finance literature largely hinge on the above two issues, they warrant careful analysis. For this reason, two essays in this dissertation are used to address the two issues. The first essay uses the bootstrap method to investigate the explanatory ability of the GLPM. The testing framework used in this essay has a high testing standard emphasizing whether there is a systematic component of asset returns left unexplained by the GLPM and places no restrictions on the sampling distribution of the statistic. The test results indicate that no more than 3% of sample stocks are mispriced according to the GLPM during any given five-year test period in the 1948-1997 sample period.

Based on the testing results built in the first essay, the second essay analyzes the long-term performance evaluation framework based on the GLPM. This study is in the line with the arguments of Loughran and Ritter (2000) that long-term performance evaluation requires a normative (equilibrium) asset pricing model. In addition, this evaluation framework uses the bootstrap tests on an individual stock basis to mitigate the skewness problem of abnormal return distributions and avoid the aggregation problems in many long-term performance tests. Finally, the specification of our long-term performance evaluation framework is evaluated using samples of randomly selected NYSE/AMEX/NASDAQ stocks and simulated random event dates. Our simulation results show that the long-term performance evaluation framework based on the GLPM is well-specified.

CHAPTER 2

EMPIRICAL ASSET PRICING IN AN ALTERNATIVE FUNCTIONAL FORM

2.1 Introduction

The message from Fama and French (1992, 1993, 1996) is loud and clear. The Capital Asset Pricing Model (CAPM) of Sharpe (1964), Lintner (1965), and Mossin (1966) is not capable of empirically explaining asset returns. Instead, Fama and French propose size and book-to-market ratio, along with the market portfolio, as the risk factors that generate asset expected returns. However, there is a controversy over why size and book-to-market ratio are relevant factors. For one thing, these two characteristics are not derived from an equilibrium model. Hence, one may question the *ad hoc* nature of the Fama-French three factor model.[3] As more and more studies indicate that characteristics with no exposure to underlying economic risk factors could be related to asset returns (e.g., Berk (1995), Daniel and Titman (1997)) and distress appears to have no systematic risk component (Dichev, 1998), the urge to empirically show that asset returns follow a well-specified theoretical model seems to grow.[4] In response to this sentiment, this study proposes a new test, which allows us to study asset returns in a less restricted empirical framework, to examine a new equilibrium model that is, on theoretical grounds, more

[3] A related argument is that the relationship between characteristics and asset returns can be due to data snooping, e.g., MacKinlay (1995).

[4] The robustness of size and book-to-market ratio as pervasive factors is also challenged. Knez and Ready (1997) find that size loses its explanatory power when the one percent most extreme observations are trimmed each month. Loughran (1997) documents that book-to-market ratio is not a determinant of asset returns outside of January for large firms.

robust than the CAPM. The test results indicate that empirical asset pricing is largely consistent with the new model.

The CAPM has traditionally served as the backbone of empirical asset pricing. Despite its popularity, the mean-variance preferences underlying the CAPM have some theoretical drawbacks. For example, Dybvig and Ingersoll (1982) and Jarrow and Madan (1997) illustrate that, under mean-variance preferences, arbitrage opportunities can exist in markets trading options. Artzner, Delbaen, Eber, and Heath (1997) demonstrate that investors with mean-variance preferences may reject a free lottery ticket. Leland (1999) shows that if market portfolio returns are independently and identically distributed, then the market portfolio is mean-variance inefficient. Lin (1999b) demonstrates that, under mean-variance pricing, investors in equilibrium can increase the values of their portfolios by giving up ex ante cash flows in states where the market portfolio has a sufficiently large return. These drawbacks suggest that the CAPM has limitations and could be invalid under no restriction on the upside potential.

Empirical evidence on risk aversion also suggests that the variance of portfolio returns does not fully characterize economic agents' perception of and behavior toward risk. A number of studies of risky choice involving monetary outcomes have documented that economic agents are averse to losses. For example, Kahneman and Tversky (1979) and Tversky and Kahneman (1986, 1991, 1992) find that economic agents are much more sensitive to losses than to gains. This loss-averse view is also advocated by Fishburn (1977) and Benartzi and Thaler (1995).

Recently, Lin (1999a) has shown that loss aversion can be incorporated into equilibrium asset pricing. Under the assumption that investors become more averse to losses when they expect to lose more, Lin derives an equilibrium asset pricing model, the Gain-Loss Pricing Model (GLPM). This model has two theoretical advantages over the CAPM. First, it does not have the above theoretical drawbacks of the CAPM. Second, it is valid for all financial assets at the Pareto-optimal allocation of risk. Furthermore, the GLPM is mathematically equivalent to the Mean-First Lower Partial Moment (MLPM(1)) model of Bawa and Lindenberg (1977) and hence incorporates the intuitive appeal of downside risk measures.[5]

Since the GLPM has economic intuition and robustness, a question naturally arises: can it empirically explain asset returns? To answer this question, we propose a new testing framework and use the bootstrap method to examine whether the pricing of U.S. common stocks is consistent with the GLPM. With the use of the bootstrap confidence intervals, this testing framework allows asset pricing tests to be conducted on an individual asset or portfolio basis and places no distributional restrictions on the sampling distribution of the statistic. Because tests are applied to each sample asset, this testing framework allows us to address an empirical question: what percentage of common stocks are priced according to the theory? This question has practical and empirical implications to the study of asset pricing, performance evaluation, and investment strategies because the smaller the number of

[5] Since the GLPM is mathematically equivalent to the MLPM(1) model, the empirical results obtained in this study also support the MLPM(1) model.

assets mispriced according to the theory, the less feasible it is to exploit mispricing opportunities.

Our test results show that, during the 1948-1997 sample period, the maximum rejection rates of the null hypothesis of no mispricing in the 10 five-year test periods are 3.07% and 8.18% for the 1% and 5% levels of significance, respectively. Given the allowable level of the theoretical type-I errors and the randomness in the level of the type-I errors, the results imply that no more than 3% of sample stocks are mispriced in any of the five-year test periods. The mean rejection rates of the null hypothesis under the GLPM, 2.03% and 7.04%, are only about 1% and 2% away from the 1% and 5% levels of significance, respectively. In addition, the rejection of the null hypothesis is not severe in small size portfolios. Overall, it appears that the GLPM does a good job of explaining stock returns in terms of their risk-return relation.

The remainder of this essay is organized as follows. Section 2.2 introduces the GLPM and its relationship with the MLPM(1) model. Section 2.3 describes the data. Section 2.4 illustrates the testing framework and the bootstrap method. Section 2.5 provides empirical results. Section 2.6 concludes this essay.

2.2 The Gain-Loss Pricing Model (GLPM)

This section first introduces the GLPM. We then discuss its equivalence to the MLPM(1) model.

2.2.1 The GLPM

As Greenblatt (1997) intuitively describes, "comparing the risk of loss in an investment to the potential gain is what investing is all about." Based on this

intuition, Lin (1999a) proposes that investors have gain-loss utility functions with the form:

$$U(G_{ps}, L_{ps}) = W_0\, G_{ps} - [1 + a/2\, E(L_{ps})]\, W_0\, L_{ps} \qquad (2.1)$$

where

$G_{ps} = (R_{ps}-R_f)$ if $R_{ps} > R_f$, and $G_{ps} = 0$ otherwise;

$L_{ps} = -(R_{ps}-R_f)$ if $R_{ps} < R_f$, and $L_{ps} = 0$ otherwise;

R_{ps} is the return on portfolio P in state s; W_0 denotes initial wealth; R_f is the risk-free rate; and a is the loss aversion coefficient with $a > 0$. Investors may have different loss aversion coefficients.

The gain-loss utility functional form in (2.1) is bilinear, which has an empirical basis. In their experimental studies, Kahneman and Tversky (1979), Kahneman, Knetsch, and Thaler (1990), and Tversky and Kahneman (1986, 1991, 1992) find that individuals making decisions under conditions with uncertain monetary outcomes exhibit approximately a bilinear gain-loss utility function. The utility function in (2.1) further assumes that investors become more averse to losses when they expect to lose more in the investment.[6] With a single parameter, a, for loss aversion, this form of gain-loss utility is designed to reflect the risk of loss as the primary concern of investors who, by taking risk, expect to be rewarded with potential gains from investing.

[6] Under this bilinear utility function, an investor's loss aversion increases when she expects to lose more from investing, which is intuitively appealing. In contrast, the typical bilinear utility function assumed in this literature has a constant loss aversion, e.g., $U(R_{ps}, L_{ps}) = (R_{ps}-R_f) - aL_{ps}$ (Sharpe, 1998).

Lin (1999a) shows that, in a frictionless, competitive economy, the optimal investment decision for any investor with the gain-loss utility function involves: (i) constructing the best risky portfolio with a ratio of expected gain to expected loss that is the highest among those that can be constructed in the economy, (ii) investing in this best risky portfolio until the investor's marginal rate of substitution of expected gain for expected loss is equal to the ratio of the expected gain to the expected loss of the best risky portfolio, and (iii) lending the remaining funds, or borrowing if insufficient, at the risk-free rate.

He further shows that if investors have homogeneous expectations, then, in equilibrium, the market portfolio, M, must be the best risky portfolio with the highest gain-loss ratio and any asset *i* must be priced such that:

$$\frac{E(MRG_i)}{E(MRL_i)} = \frac{E(G_m)}{E(L_m)} \tag{2.2}$$

where

$MRG_i = (R_i - R_f)$ if $R_m > R_f$ and $MRG_i = 0$ otherwise;

$MRL_i = -(R_i - R_f)$ if $R_m < R_f$, and $MRL_i = 0$ otherwise;

$G_m = (R_m - R_f)$ if $R_m > R_f$, $G_m = 0$ otherwise; and

$L_m = -(R_m - R_f)$ if $R_m < R_f$, $L_m = 0$ otherwise.

Equation (2.2) is the gain-loss pricing model (GLPM), which postulates that, in equilibrium, the ratio of expected market-related gain to expected market-related loss of asset *i*, $E(MRG_i)/E(MRL_i)$, is equal to the ratio of expected gain to expected loss of the market portfolio, $E(G_m)/E(L_m)$. Since the gain-loss ratio of the market

portfolio, $\pi \equiv E(G_m)/E(L_m)$, is invariant to all assets, the GLPM predicts that the higher the expected market-related loss of an asset, the higher the expected market-related gain investors requires to compensate for the loss. Thus, under the GLPM, the gain-loss ratio of the market portfolio defines the gain-loss tradeoff for individual assets.

On theoretical grounds, the GLPM is superior to the CAPM in two respects. First, Lin (1999a) shows that the GLPM results in a positive pricing operator. Consequently, the GLPM eliminates the arbitrage opportunities arising from the possibility of negative prices for call options in the CAPM (Dybvig and Ingersoll (1982), Jarrow and Madan (1997)). Second, Lin demonstrates that the GLPM holds before and after the introduction of derivative assets that complete the market. In contrast, Dybvig and Ingersoll (1982) show that if enough derivative assets are added to complete the market, the CAPM would collapse. Given the theoretical advantages, the focus of this study is to see how well the GLPM explains empirical asset returns.

2.2.2 The Relationship between the GLPM and the MLPM(1) Model

Another advantage of the GLPM is that it captures the appeal of downside risk measures. To many, a downside risk measure seems more plausible than variance as a measure of risk because of its consistency with the way economic agents actually perceive risk (Markowitz (1959), Mao (1970)). In the following proposition we prove the mathematical equivalence between the GLPM and the Mean-first order Low Partial Moment, MLPM(1), model of Bawa and Lindenberg (1977).

Proposition: The GLPM is mathematically equivalent to the MLPM(1) Model.

Proof:

The family of the MLPM models of Hogan and Warren (1974) and Bawa and Lindenberg (1977) can be specified as:

$$E(R_i) = R_f + \beta_i^{MLPM(n)} [E(R_m) - R_f] \tag{2.3}$$

$$\beta_i^{MLPM(n)} = \frac{\int_{R_m=-\infty}^{R_f} \int_{R_i=-\infty}^{\infty} (R_f - R_m)^{n-1} (R_f - R_i) dF(R_i, R_m)}{\int_{-\infty}^{R_f} (R_f - R_m)^n \, dF(R_m)}$$

where $E(R_m)$ is the expected return of the market portfolio, $E(R_i)$ is the expected return of asset i, and $n = 1$ or 2. When $n = 1$, equation (2.3) is the MLPM(1) model; when $n = 2$, equation (2.3) is the MLPM(2) model. The following shows that the GLPM and MLPM(1) model are mathematically equivalent.

$$E(MRG_i) / E(MRL_i) = E(G_m) / E(L_m)$$

$$\Leftrightarrow (E(MRG_i) / E(MRL_i)) - 1 = (E(G_m) / E(L_m)) - 1$$

$$\Leftrightarrow E(MRG_i) - E(MRL_i) = ((E(G_m) / E(L_m)) - 1) E(MRL_i) \tag{2.4}$$

$$\Leftrightarrow E(R_i) - R_f = ((E(G_m) / E(L_m)) - 1) E(MRL_i)$$

$$\Leftrightarrow E(R_i) = R_f + (E(MRL_i) / E(L_m)) (E(G_m) - E(L_m))$$

$$\Leftrightarrow E(R_i) = R_f + (E(MRL_i) / E(L_m)) (E(R_m) - R_f) \tag{2.5}$$

$$\Leftrightarrow E(R_i) = R_f + \beta_i^{MLPM(1)} (E(R_m) - R_f)$$

Q.E.D.

Since the GLPM is mathematically equivalent to the MLPM(1), it can capture the appeal of the downside risk measure in the MLPM(1) model. However, as will be shown, our empirical testing framework emphasizes the gain-loss tradeoff as specified in the GLPM, instead of following the traditional approach of the return-beta relation as in the CAPM and MLPM(1) model.

2.3 The Sample

The monthly returns data are obtained from the 1997 Center for Research in Security Prices (CRSP) monthly files. The data comprises the set of all NYSE, AMEX, and NASDAQ stocks from January 1948 to December 1997. The CRSP value weighted returns series is used to proxy for the market returns. We divide the 50-year sample period into 10 non-overlapping test periods, each consisting of five years.

The data set is adjusted for the delisting bias to mitigate survivorship bias. Shumway (1997) finds that most of the missing delisting returns in the CRSP tapes are associated with negative events and suggests a −30% delisting monthly return for NYSE and AMEX stocks. Similarly, Shumway and Warther (1998) suggest a corrected delisting return of −55% for NASDAQ stocks. Following Shumway (1997) and Shumway and Warther (1998), we classify delisting codes 500 and 505 through 588 as negative-performance-related and adjust the data set for the missing delisting returns.

Since, in general, there is no return observations available after the delisting month, some delisting stocks may not have enough observations to construct

bootstrap samples. To be included in the analysis, a sample stock must have at least 24 consecutive monthly return observations in a given five-year test period.

2.4 The Testing Framework

In this section, the testing design and its strengths are discussed. The construction of confidence intervals based on the bootstrap percentiles is introduced. Hypothesis testing with these empirical confidence intervals is then outlined.

2.4.1 The Testing Design

The theory in equation (2.2) indicates that the pricing error for asset i, defined as $PE_i = E(MRG_i)\,E(L_m) - E(MRL_i)\,E(G_m)$, should be zero if the pricing of the asset is consistent with the GLPM. Hence, we can test the GLPM by testing the null hypothesis H_0: $PE_i = 0$. However, PE_i is a function of the unknown distributions of MRG_i, L_m, MRL_i, and G_m. Correspondingly, we propose a sample pricing error, $\widehat{PE}_i$, as:

$$\widehat{PE}_i = \hat{E}(MRG_i)\,\hat{E}(L_m) - \hat{E}(MRL_i)\,\hat{E}(G_m) \tag{2.6}$$

where $\hat{E}(MRG_i)$, $\hat{E}(MRL_i)$, $\hat{E}(G_m)$, and $\hat{E}(L_m)$ are the sample mean estimates of $E(MRG_i)$, $E(MRL_i)$, $E(G_m)$, $E(L_m)$. Given that, for a given asset i, each return observation is a random draw from a stable distribution that characterizes the asset, the sample mean estimates are the unbiased estimates of the true variables. That is,

$$\hat{E}(MRG_i) = E(MRG_i) + \xi_i$$

$$\hat{E}(MRL_i) = E(MRL_i) + \delta_i$$

$$\hat{E}(G_m) = E(G_m) + \varepsilon_m$$

$$\hat{E}(L_m) = E(L_m) + \zeta_m$$

and the estimation errors, ξ_t, δ_t, ε_m, and ζ_m, are distributed with zero means. As a result, the sample pricing error $\hat{PE}_t$ is an unbiased estimate of the pricing error PE_t.

The testing framework based on PE_t and $\hat{PE}_t$ has two empirical strengths. First, the tests based on the GLPM impose a stringent standard. In this study, the null hypothesis is H_0: $PE_t = 0$. Similar to the intercept term test of Black, Jensen, and Scholes (1972), this test focuses on whether there is a systematic component of asset returns left unexplained by the model.[7] Consequently, it imposes a higher testing standard than the two-stage cross-sectional regressions of Fama and MacBeth (1973), in which the null hypothesis is whether the systematic risk measure is related to asset return.

Second, testing H_0: $PE_t = 0$ can be conducted for each sample security. This advantage allows us to answer the question: what proportion of individual assets are priced according to the GLPM? This is an important question because if only a small number of assets are mispriced, the feasibility of exploiting mispricing opportunities will be considerably limited.

Asset pricing tests based on individual assets also avoid the problematic procedure of sorting securities into groups. For example, Litzenberger and Ramaswamy (1979) argue that sorting procedures reduce the information content of financial data. Lo and MacKinlay (1990) demonstrate that asset pricing tests are subject to data-snooping bias when sorting procedures are based on empirical-based

[7] To see the structural similarity between the two tests, rearrange the regression specification of Black, Jensen, and Scholes (1972) to get: $a + \varepsilon(t) = R(t) - R_f(t) - b\,[R_m(t) - R_f(t)]$. Since $\varepsilon(t)$ is assumed to be mean zero in a LS regression, the intercept term test essentially tests whether the pricing errors under the CAPM systematically differs from zero. Note that the RHS of the above equation and the RHS of equation (2.6) are both the sample counterparts to the theoretical models.

variables. Berk (2000) points out that if sorting variables are identified within the sample, part of the variations within or between groups might be spurious. Consequently, he argues the true asset pricing model can be shown to have no explanatory power when sorting procedures are employed.

2.4.2 The Percentile Intervals

This study uses the nonparametric bootstrap method. Suppose we are given a random sample $X_i = (x_{i1}, x_{i2}, \ldots, x_{iT})'$ for asset i, where x_{it}, $t = 1, 2, \ldots, T$, is an 1×4 vector of observations (MRG_{it}, MRL_{it}, G_{mt}, L_{mt}) from an unknown probability distribution F.[8] The purpose of the bootstrap method is to estimate the sampling distribution of the parameter of interest $PE_i = E(MRG_i)\,E(L_m) - E(MRL_i)\,E(G_m)$, on the basis of the given sample.

Following Efron (1979), let $\hat{F}$ be the empirical probability distribution that puts probability $1/T_i$ on each value x_{it}, $t = 1, 2, \ldots, T_i$. The bootstrap method is to estimate the sampling distribution of PE_i by the bootstrap distribution of $\hat{PE}_i{}^* = \hat{E}(MRG_i{}^*)\,\hat{E}(L_m{}^*) - \hat{E}(MRL_i{}^*)\,\hat{E}(G_m{}^*)$. Each $X_i{}^* = (x_{i1}{}^*, x_{i2}{}^*, \ldots, x_{iT_i}{}^*)'$ is a random sample of size T_i drawn with replacement from the population of T_i objects $(x_{i1}, x_{i2}, \ldots, x_{iT_i})'$, with $\hat{F}$ held fixed at its observed value.[9] The calculation of $\hat{PE}_i{}^*$ follows exactly the same functional form as $\hat{PE}_i$.[10] To obtain the empirical bootstrap

---

[8] The symbol ' denotes the transpose.

[9] In light of time dependence in the return data, we also use block resampling. Using Hall, Horowitz, and Jing's (1995) optimal asymptotic formula for block length, we use a block size of two for the bootstrap tests. The results are similar to the baseline results that we report.

[10] For example, suppose that $T_i = 60$. $\hat{E}(MRG_i{}^*)$ then may be obtained by computing $(MRG_{i3} + MRG_{i19} + MRG_{i3} + MRG_{i58} + \ldots + MRG_{i24})/60$.

distribution of $\hat{PE}_t$*, Monte Carlo realizations of X_t* are generated 1,000 times to obtain 1,000 $\hat{PE}_t$*s in this study.[11]

2.4.3 Hypothesis Testing

Let H be the empirical bootstrap distribution of $\hat{PE}_t$*. The $1-2r$ percentile two-tailed confidence interval is defined by the r and $1-r$ percentiles of H. For example, at the 5% level of significance, r is 2.5% and the lower and upper bounds of the 95% confidence intervals are the 25th and the 975th observations of the ranked $\hat{PE}_t$* in our bootstrap setting. Similarly, at the 1% level of significance, the lower and upper bounds of the 99% confidence intervals are the 5th and the 995th observations of the ranked $\hat{PE}_t$*. Once the confidence intervals are obtained, statistical inferences can be made in the usual way. If the $1-2r$ percentile confidence interval covers the hypothesized value of zero, the null hypothesis H_0: $PE_t = 0$ is accepted; if the interval does not cover zero, the null hypothesis is rejected. Similarly, critical values can be obtained by identifying the $2r$ or $1-2r$ percentiles of H for one-tailed tests.

We use the percentile intervals based on the nonparametric bootstrap because, as the results of the normality test on the bootstrap distribution of $\hat{PE}_t$* in the next section indicate, the sampling distribution of PE_t is not normal.[12] The nonparametric bootstrap is free from distributional assumptions; it only requires a random sample and a proposed statistic of PE_t. This property is particularly

[11] According to Efron and Tibshirani (1993), the number of bootstrap replications typically ranges from 50 to 200.

[12] PE_t is the difference of two random products. Even we are willing to assume the four random variates are normally distributed, the resulting random products and their difference will not be normally distributed. Therefore, it does not appear to be feasible to derive an analytical test statistic for PE_t.

appealing for this study since the distribution F is unknown and may not belong to a class of analytically tractable distributions. The strength of the bootstrap method is that, even if the distribution F is unknown, this method leads to a consistent estimator of PE_i (Efron, 1979).

2.5 Empirical Results

This section further illustrates the testing framework based on the null hypothesis, H_0: $PE_i = 0$. The bootstrap results are then reported. As a comparison, the same bootstrap testing procedures are also applied to examine the CAPM. We then extend the comparison to the case where artificial returns are added to return observations. Finally, the distribution of the rejection of the null hypothesis across sizes is examined to further demonstrate the empirical power of the GLPM.

2.5.1 Estimating the Sampling Distributions of the Parameter PE_i

The mean values of $\hat{E}(MRG_i)$, $\hat{E}(MRL_i)$, $\hat{E}(G_m)$, and $\hat{E}(L_m)$ in each five-year test period are reported in table 2.1. During the 1948-1997 sample period, the average monthly expected market related gain, $\hat{E}(MRG_i)$, has a mean value of 2.24%, the average monthly expected market related loss, $\hat{E}(MRL_i)$, has a mean value of 1.24%. In the same period, the time-series mean value of the average monthly expected gain on the market index, $\hat{E}(G_m)$, is 1.96%, and the time-series mean value of the average monthly expected loss on the market index, $E(L_m)$, is 1.19%, suggesting that the estimate of the gain-loss ratio of market index, $\hat{\pi}$, is 1.65 (1.96%/1.19%) during the 1948-1997 sample period.

For each sample stock in each of the 10 five-year test periods from 1948 to 1997, Fisher's (1930) cumulant test for normality is applied to the bootstrap

Table 2.1

Average Estimates of $\hat{E}(MRG_i)$, $\hat{E}(MRL_i)$, $\hat{E}(G_m)$, and $\hat{E}(L_m)$, 1948-1997

This table presents the sample estimates of $E(G_m)$. and $E(L_m)$ and the mean values of the sample estimates of $E(MRG_i)$ and $E(MRL_i)$ in the ten non-overlapping five-year test periods during the 1948-1997 sample period. MRG_i, MRL_i, G_m, and L_m is a random vector of the market-related gain of asset i, a random vector of the market-related loss of asset i, a random vector of the gain on the market index, and a random vector of the loss on the market index, respectively, in a test period. Each element of MRG_i is defined as $MRG_{it} = (R_{it} - R_{ft})$ if $R_{mt} > R_{ft}$ and $MRG_{it} = 0$ otherwise; each element of MRL_i is defined as $MRL_{it} = -(R_{it} - R_{ft})$ if $R_{mt} < R_{ft}$, and $MRL_{it} = 0$ otherwise; each element of G_m is defined as $G_{mt} = (R_{mt} - R_{ft})$ if $R_{mt} > R_{ft}$, $G_{mt} = 0$ otherwise; and each element of L_m is defined as $L_{mt} = -(R_{mt} - R_{ft})$ if $R_{mt} < R_{ft}$, $L_{mt} = 0$ otherwise. The data comprises the set of all NYSE, AMEX, and NASDAQ stocks with a minimum of two years' continuous return observations. The CRSP value weighted returns series represents the market returns. Following Shumway (1997) and Shumway and Warther (1998), the data set is adjusted for the delisting bias.

Test Period	# of Sample Stocks	Average $\hat{E}(MRG_i)$	Average $\hat{E}(MRL_i)$	$\hat{E}(G_m)$	$\hat{E}(L_m)$
1948-1952	939	0.0237	0.0117	0.0227	0.0100
		(0.0083)	(0.0054)	(0.0030)	(0.0024)
1953-1957	1010	0.0165	0.0093	0.0193	0.0102
		(0.0074)	(0.0056)	(0.0030)	(0.0022)
1958-1962	1015	0.0223	0.0119	0.0204	0.0110
		(0.0090)	(0.0058)	(0.0028)	(0.0029)
1963-1967	1728	0.0295	0.0077	0.0155	0.0077
		(0.0174)	(0.0094)	(0.0023)	(0.0021)
1968-1972	1981	0.0212	0.0215	0.0184	0.0164
		(0.0138)	(0.0100)	(0.0031)	(0.0033)
1973-1977	4155	0.0230	0.0106	0.0176	0.0216
		(0.0175)	(0.0174)	(0.0044)	(0.0038)
1978-1982	4119	0.0304	0.0161	0.0226	0.0178
		(0.0154)	(0.0111)	(0.0038)	(0.0038)
1983-1987	4496	0.0199	0.0212	0.0219	0.0153
		(0.0187)	(0.0173)	(0.0037)	(0.0045)
1988-1992	4911	0.0184	0.0084	0.0190	0.0112
		(0.0205)	(0.0150)	(0.0031)	(0.0026)
1993-1997	4414	0.0186	0.0059	0.0189	0.0076
		(0.0170)	(0.0092)	(0.0024)	(0.0019)
Mean		0.0224	0.0124	0.0196	0.0119
		(0.0145)	(0.0106)	(0.0032)	(0.0030)

distribution of the 1,000 $\hat{PE}_i$ * s.[13] Table 2.2 reports the results of this test. It is clear that the bootstrap distribution of $\hat{PE}_i$ * is far from normal for a large percentage of sample stocks. During the 50 year sample period, 52.61% and 36.69% of sample stocks exhibit significant skewness and kurtosis, respectively, in the bootstrap distribution of PE_i *. The combined test of skewness and kurtosis also presents similar results, an average rejection rate of 57.13%. These average rejection rates are all far away from the 5% level of significance.

Simulations are used to numerically show that the bootstrap distribution of the 1,000 PE_i * s will be normal if the sampling distribution of PE_i is normal in our bootstrap setting. First, a 60 × 1 of unit normal random vector of PE_i is drawn for each artificial stock.[14] This procedure is repeated to generate 10 sets of artificial stocks. The numbers of artificial stocks in the 10 sets of data correspond to the number of sample stocks in the 10 test periods from 1948 to 1997. Monte Carlo realizations of bootstrap samples are then independently generated to obtain 1,000 PE_i * s for each artificial stock. Applying Fisher's cumulant test for normality to the 10 sets of simulations shows that the bootstrap distribution of $\hat{PE}_i$ * are quite close to normal. In table 2.3, the rejection rates for the skewness test, kurtosis test, and the combined test are 7.78%, 4.57%, and 6.40%, respectively. These rejection rates are all quite close to the 5% level of significance of the test. Based on these results,

[13] The test statistic for skewness is $u1=(k3/(k2)^{3/2}) \times (n/6)^{1/2}$. The test statistic for kurtosis is $u1= (k4/(k2)^2) \times (n/24)^{1/2}$. The test statistics for the combined test is $x2=u1^2 + u2^2$. $k1=m1/n$. $k2=(n\times m2- m1^2)/(n(n-1)$. $k3=((n^2\times m3)-(3\times n\times m2\times m1)+(2\times m1^3))/(n\times(n-1)\times(n-2))$. $k4=((n^3-n^2)\times m4-4\times(n^2+n)\times m3\times m1-3\times(n^2- n)\times m2^2+12\times m2\times m1^2-6\times m1^4)/(n\times(n-1)\times(n-2)\times(n-3))$. m1. m2. m3. and m4 are the sample 1st. 2nd. 3rd. and 4th moments. respectively. The critical value for the skewness and kurtosis statistics are 1.96. The critical value for the combined test is 5.99.

[14] The results are invariant to the scaling of normal random vectors.

Table 2.2

Tests of Normality on the Bootstrap Distributions of $\hat{PE}_i$* for Each Sample Stock, 1948-1997

This table presents testing results of the Fisher's cumulant test for normality on the bootstrap distribution of sample pricing errors for each sample stock. The data comprises the set of all NYSE, AMEX, and NASDAQ stocks with a minimum of two years' continuous return observations. Following Shumway (1997) and Shumway and Warther (1998), the data set is adjusted for the delisting bias. Monte Carlo realizations of bootstrap samples are independently generated to obtain 1,000 bootstrap pricing errors for each sample stock. Fisher's cumulative test is then applied to the bootstrap pricing errors for each sample stock. The critical value for skewness statistic and kurtosis statistic is 1.96 at the 5% level of significance. The critical value for the combined test is 5.99 at the 5% level of significance.

		Skewness		Kurtosis		Combined Test	
Test Period	# of Sample Stocks	# (and %) of Stocks Rejected	Test Statistic	# (and %) of Stocks Rejected	Test Statistic	# (and %) of Stocks Rejected	Test Statistic
1948-1952	939	380 (40.47%)	0.72	202 (21.51%)	0.96	403 (42.92%)	9.79
1953-1957	1010	378 (37.43%)	0.85	178 (17.62%)	0.82	386 (38.22%)	9.18
1958-1962	1015	503 (49.56%)	1.59	304 (29.95%)	1.47	545 (53.69%)	16.22
1963-1967	1728	1049 (60.71%)	2.42	800 (46.30%)	2.48	1155 (66.84%)	38.43
1968-1972	1981	889 (44.88%)	1.14	548 (27.66%)	1.39	934 (47.15%)	19.16
1973-1977	4155	2989 (71.94%)	3.28	2314 (55.69%)	3.05	3242 (78.03%)	54.84
1978-1982	4119	1853 (44.99%)	1.30	1297 (31.49%)	1.49	2014 (48.90%)	19.55
1983-1987	4496	2744 (61.03%)	0.55	2497 (55.54%)	2.99	3187 (70.89%)	50.99
1988-1992	4911	2583 (52.60%)	1.39	1837 (37.41%)	1.93	2785 (56.71%)	29.01
1993-1997	4414	2763 (62.60%)	1.28	1929 (43.70%)	2.21	3001 (67.99%)	42.62
Mean		(52.61%)	1.45	(36.69%)	1.88	(57.13%)	28.98

Table 2.3

Tests of Normality on the Bootstrap Distributions of Artificial $\hat{PE_i}$* from Unit Normal Random Samples

This table presents testing results of the Fisher's cumulant test for normality on artificial pricing errors. A 60 × 1 of unit normal random vector of PE_i is first drawn for each artificial stock. This procedure is repeated to generate 10 sets of artificial stocks. The numbers of artificial stocks in the 10 sets are 939, 1010, 1015, 1728, 1981, 4155, 4119, 4496, 4911, and 4414. Monte Carlo realizations of bootstrap samples are independently generated to obtain 1,000 bootstrap pricing errors for each artificial stock. Fisher's cumulative test is then applied to the bootstrap pricing errors for each artificial stock. The critical value for skewness statistic and kurtosis statistic is 1.96 at the 5% level of significance. The critical value for the combined test is 5.99 at the 5% level of significance.

		Skewness		Kurtosis		Combined Test	
Artificial Simulation	# of Artificial Stocks	# (and %) of Stocks Rejected	Test Statistic	# (and %) of Stocks Rejected	Test Statistic	# (and %) of Stocks Rejected	Test Statistic
# 1	939	65 (6.92%)	0.04	34 (3.62%)	-0.13	53 (5.64%)	2.09
# 2	1010	96 (9.50%)	0.01	44 (4.36%)	-0.04	75 (7.43%)	2.33
# 3	1015	87 (8.57%)	0.04	41 (4.03%)	-0.09	55 (5.42%)	2.21
# 4	1728	105 (6.08%)	-0.06	89 (5.15%)	-0.08	96 (5.56%)	2.12
# 5	1981	170 (8.58%)	0.01	90 (4.54%)	-0.06	139 (7.02%)	2.29
# 6	4155	332 (7.99%)	0.02	205 (4.93%)	-0.06	286 (6.88%)	2.25
# 7	4119	329 (7.99%)	0.02	205 (4.98%)	-0.06	286 (6.94%)	2.26
# 8	4496	319 (7.10%)	-0.02	204 (4.54%)	-0.05	269 (5.98%)	2.20
# 9	4911	368 (7.49%)	-0.02	246 (5.01%)	-0.05	334 (6.80%)	2.24
# 10	4414	335 (7.59%)	0.02	198 (4.49%)	-0.05	277 (6.28%)	2.21
Mean		(7.78%)	0.01	(4.57%)	-0.07	(6.40%)	2.22

it appears that the bootstrap method can approximate the distribution of the PE_i and that the sampling distributions of PE_i for majority of the sample stocks are not normally distributed.

2.5.2 Test Results of the Null Hypothesis H_0: $PE_i = 0$

The bootstrap method is used to construct the empirical distribution of $\hat{PE}_i*$, $\hat{H}$. The 95% and 99% confidence intervals are defined by the 2.5 and 97.5 percentiles and 0.5 and 99.5 percentiles, respectively, of $\hat{H}$. If the confidence interval covers the hypothesized value of zero, then the null hypothesis H_0: $PE_i = 0$ is accepted; if the interval does not cover zero, then the null hypothesis is rejected.

Before using this testing framework on sample stocks, the bootstrap method is numerically verified by applying it to the 10 sets of artificial stocks created earlier. These artificial data sets are actually generated in an ideal pricing environment in which pricing error is drawn from a unit normal distribution with mean zero. Consequently, if the bootstrap method is well-specified, the empirical rejection rates of these artificial data sets should be close to the levels of significance. Table 2.4 reports the test results. The rejection rates for the 10 sets of artificial stocks range from 1.31% to 1.98% for the 1% level of significance with a mean rejection rate of 1.60%. At the 5% level of significance, the rejection rates range from 5.30% to 6.90% with an average of 6.08%. As expected, these rejection rates are close to the 1% and 5% levels of significance.

Note that the cross-sectional rejection frequency will be close to, but not necessarily equal to, the level of significance when the number of sample stocks is a

Table 2.4

Testing Results of the Null Hypotheses H_0: $PE_i = 0$ with Bootstrap Confidence Intervals on Each Artificial Stock

This table presents testing results of the null hypothesis H_0: $PE_i = 0$ with bootstrap confidence intervals on each artificial stock. A 60×1 of unit normal random vector of PE_i is first drawn for each artificial stock. This procedure is repeated to generate 10 sets of artificial stocks. The numbers of artificial stocks in the 10 sets are 939, 1010, 1015, 1728, 1981, 4155, 4119, 4496, 4911, and 4414. Monte Carlo realizations of bootstrap samples are independently generated to obtain 1,000 bootstrap pricing errors for each artificial stock. The empirical distribution of these bootstrap pricing errors is then used to construct the bootstrap confidence intervals. The 95% and 99% percentile confidence intervals used in this study are defined by the 2.5 and 97.5 percentiles and the 0.5 and 99.5 percentiles, respectively, of the distribution of the bootstrap pricing errors. If the confidence interval covers the hypothesized value of zero, then the null hypothesis will be accepted; if the interval does not cover zero, then the null hypothesis will be rejected. The 95% confidence interval for the rejection rate in the last column is defined as $a \pm 2\times((a \times (1-a))/N)^{1/2}$, where N is the number of artificial stocks and a is the level of significance used in the bootstrap tests.

Artificial Simulation	# of Artificial Stocks	Bootstrap Distribution (%) 0.5	99.5	Total	95% Confidence Interval
Panel A: 1% Level of Significance					
		Rejection Rate (%)			
# 1	939	0.53	1.06	1.59	0.35 ~ 1.65
# 2	1010	0.59	1.39	1.98	0.37 ~ 1.63
# 3	1015	0.89	1.08	1.97	0.38 ~ 1.62
# 4	1728	0.46	1.04	1.50	0.52 ~ 1.48
# 5	1981	0.66	0.76	1.42	0.55 ~ 1.45
# 6	4155	0.55	1.08	1.63	0.69 ~ 1.31
# 7	4119	0.83	0.97	1.80	0.69 ~ 1.31
# 8	4496	0.58	0.73	1.31	0.70 ~ 1.30
# 9	4911	0.79	0.73	1.52	0.72 ~ 1.28
# 10	4414	0.63	0.70	1.33	0.70 ~ 1.30
Mean		0.65	0.95	1.60	
		Bootstrap Distribution (%)			
		2.5	97.5	Total	
Panel B: 5% Level of Significance					
# 1	939	2.34	3.30	5.60	3.58 ~ 6.42
# 2	1010	3.27	2.97	6.24	3.63 ~ 6.37
# 3	1015	3.25	3.65	6.90	3.63 ~ 6.37
# 4	1728	3.41	2.43	5.84	3.95 ~ 6.05
# 5	1981	2.52	3.08	5.60	4.02 ~ 5.98
# 6	4155	3.06	3.56	6.62	4.32 ~ 5.68
# 7	4119	3.11	3.35	6.46	4.32 ~ 5.68
# 8	4496	2.82	2.96	5.78	4.35 ~ 5.65
# 9	4911	3.42	3.03	6.45	4.38 ~ 5.62
# 10	4414	2.54	2.76	5.30	4.34 ~ 5.66
Mean		2.97	3.11	6.08	

fairly large finite number. According to Brown and Warner (1980), suppose that the testing outcomes for each of the N assets in a test period are independent. Then at the a level of significance, the cross-sectional rejection frequency of the N assets for such a Bernoulli process has a mean of a and a standard deviation of:

$$\sigma_{Bernoulli,\,(a,\,1-a)} = \sqrt{\frac{a \times (1-a)}{N}} \tag{2.7}$$

If asset pricing exactly follows the GLPM model, we should expect that there is an approximately 95% probability that the cross-sectional rejection frequency of the N assets will fall into the 95% confidence interval of $a \pm 2 \times \sigma_{Bernoulli,\,(a,\,1-a)}$.

The 95% confidence intervals of the random rejection rates for the 10 artificial data sets are given in the last column of table 2.4. For the bootstrap tests with the 1% level of significance, eight out of the 10 rejection rates are slightly above the 95% confidence intervals of the random rejection rates. For the bootstrap tests with the 5% level of significance, five out of the 10 rejection rates are slightly above the 95% confidence intervals of the random rejection rates. The bootstrap rejection rates and the normality rejection rates for the artificial data sets are above the levels of significance because of the randomness in the original sample. The reason is that, although the original sample is drawn from a normal distribution, the original sample and its mean value can be very far away from normal and zero, respectively. Subsequently, the bootstrap samples and the bootstrap distribution that are based on the original sample is also subject to this randomness. On the other hand, as the probability that the original sample and its mean value are normally distributed and

zero, respectively, goes to one in the limit, the normality rejection rate and the bootstrap rejection rate will approach the levels of significance.

With the verification of the bootstrap method under the null hypothesis, this testing framework is applied to each sample stock in each of the 10 five-year test periods from 1948 to 1997. The testing results are reported in table 2.5. For the 10 test periods, rejection rates with the 1% level of significance range from 1.28% to 3.07% with an average of 2.03%. The mean rejection rate is 1.03% away from the 1% level of significance. With the 5% level of significance, rejection rates range from 6.01% to 8.18%. The mean rejection rate, 7.04%, is 2.04% above the 5% level of significance. Overall, given the allowable level of the theoretical type-I errors and the randomness in the level of the type-I errors, no more than 3% of sample stocks are mispriced. This evidence indicates that the pricing of most stocks is consistent with the GLPM.

We test the equality between the two-tailed rejection rates in table 2.4 and 2.5. Under the null hypothesis that the pricing of stocks is consistent with the GLPM, the bootstrap rejection rates based on sample stocks should not be statistically different from those based on artificial stocks. That is, by the Z-test for the equality between two proportions (binomial distribution), based on two bootstrap rejection rates, κ_{sample} and $\kappa_{artificial}$,

$$Z = \frac{\kappa_{sample} - \kappa_{artificial}}{\{P(1-P)\left[\frac{2}{n}\right]\}^{1/2}} \tag{2.8}$$

Table 2.5
Testing Results of the Null Hypotheses H_0: $PE_i = 0$ with Bootstrap Confidence Intervals on Each Sample Stock, 1948-1997

This table presents testing results of the null hypothesis H_0: $PE_i = 0$ with bootstrap confidence intervals on each sample stock. The data comprises the set of all NYSE, AMEX, and NASDAQ stocks with a minimum of two years' continuous return observations. Following Shumway (1997) and Shumway and Warther (1998), the data set is adjusted for the delisting bias. Monte Carlo realizations of bootstrap samples are independently generated to obtain 1,000 bootstrap pricing errors for each sample stock. The empirical distribution of these bootstrap pricing errors is then used to construct the bootstrap confidence intervals. The 95% and 99% percentile confidence intervals used in this study are defined by the 2.5 and 97.5 percentiles and the 0.5 and 99.5 percentiles, respectively, of the distribution of the bootstrap pricing errors. If the confidence interval covers the hypothesized value of zero, then the null hypothesis will be accepted; if the interval does not cover zero, then the null hypothesis will be rejected. The Z-statistics reported are based on the null hypothesis that the rejection rates between artificial and sample stocks should be equal.

		Bootstrap Distribution (%)			
Test Period	# of Sample Stocks	0.5	99.5	Total	Z-Statistic
Panel A: 1% Level of Significance					
		Rejection Rate (%)			
1948-1952	939	0.32	1.38	1.70	0.19
1953-1957	1010	0.99	2.08	3.07	1.56
1958-1962	1015	0.59	0.69	1.28	-1.21
1963-1967	1728	1.39	0.17	1.56	0.14
1968-1972	1981	1.06	0.91	1.97	1.34
1973-1977	4155	1.95	0.19	2.14	1.71
1978-1982	4119	1.53	0.41	1.94	0.47
1983-1987	4496	1.02	1.45	2.47	4.04*
1988-1992	4911	1.26	1.22	2.48	3.40*
1993-1997	4414	1.22	0.54	1.76	1.64
Mean		1.13	0.90	2.03	
		Bootstrap Distribution (%)			
		2.5	97.5	Total	
Panel A: 5% Level of Significance					
1948-1952	939	1.92	4.47	6.39	0.72
1953-1957	1010	3.27	4.46	7.73	1.31
1958-1962	1015	3.74	2.27	6.01	-0.81
1963-1967	1728	5.96	1.22	7.18	1.60
1968-1972	1981	3.13	3.33	6.46	1.14
1973-1977	4155	7.58	0.48	8.06	2.52*
1978-1982	4119	6.05	1.26	7.31	1.52
1983-1987	4496	3.80	4.38	8.18	4.47*
1988-1992	4911	3.81	2.95	6.76	0.62
1993-1997	4414	4.51	1.74	6.25	1.91
Mean		4.38	2.66	7.04	

*Significant at the 5% level of significance

where

$$P = \frac{K_{sample} + K_{artificial}}{2}$$

we should expect that the Z statistic is within ±1.96 with a 95% probability. The Z statistic is approximately normally distributed when the sample size, n, is larger than 30 (Kanji, 1993).

We find that, for bootstrap tests at the 1% and 5% levels of significance, the rejection rates of sample stocks based on the GLPM are statistically different from the rejection rates of artificial stocks in only two test periods. In most test periods, there is no significant difference between the GLPM rejection rates and the rejection rates in the ideal pricing environment. It appears that the GLPM performs quite well.

2.5.3 The Bootstrap Results Based on the CAPM

As a comparison, the above bootstrap testing procedures are applied to sample stocks to examine the CAPM. Based on the CAPM,

$$E(R_i) - R_f = \frac{\text{cov}(R_i, R_m)}{\sigma_m^{\ 2}} [E(R_m) - R_f], \tag{2.9}$$

we propose the CAPM sample pricing error, $\hat{PE}^{CAPM}{}_i$, to be:

$$PE^{CAPM}{}_i = \hat{E}(r_i)\hat{\sigma}_m{}^2 - \hat{E}(r_m)\hat{\text{cov}}(R_i, R_m) \tag{2.10}$$

where r_i is the difference vector between individual returns and the risk free rates and r_m is the difference vector between market returns and the risk free rates. Using this statistic, Monte Carlo realizations of resampled returns are generated 1,000 times to

obtain the empirical bootstrap distribution of PE^{CAPM}_i *s. If the bootstrap percentile confidence interval covers the hypothesized value of zero, the null hypothesis H_0: $PE^{CAPM}_i = 0$ is accepted; if the interval does not cover zero, the null hypothesis is rejected.

The bootstrap testing results based on the CAPM are reported in table 2.6. For the 10 test periods, rejection rates at the 1% level of significance range from 1.28% to 3.37% with an average of 2.51%. The mean rejection rate is 1.51% away from the 1% level of significance. At the 5% level of significance, rejection rates range from 6.84% to 9.20%. The mean rejection rate, 8.03%, is 3.03% above the 5% level of significance. Compared with the bootstrap rejection rates based on the GLPM, the CAPM rejection rates are 0.39% and 0.99% higher than the GLPM rejection rates for the 1% and 5% levels of significance, respectively. This indicates that the GLPM performs better than the CAPM in explaining stock returns.

The test results of the Z-test for the equality between the two-tailed rejection rates in table 2.4 and 2.6 also confirms that the GLPM outperforms the CAPM. Unlike the GLPM, which is rejected in two out of the ten test periods, the rejection rates of sample stocks based on the CAPM are statistically different from the rejection rates of artificial stocks in six and eight out of the ten test periods for bootstrap tests at the 1% and 5% levels of significance, respectively.

2.5.4 The Power of the Bootstrap Tests

The moderate bootstrap rejection rates of the GLPM could be due to either the explanatory ability of the GLPM or the lack of the power of the bootstrap test. To investigate the possibility that the bootstrap test will lead to rejection of a false

Table 2.6
Testing Results of the Null Hypotheses H_0: $PE^{CAPM}_i = 0$ with Bootstrap Confidence Intervals on Each Sample Stock, 1948-1997

This table presents testing results of the null hypothesis H_0: $PE^{CAPM}_i = 0$ with bootstrap confidence intervals on each sample stock. The data comprises the set of all NYSE, AMEX, and NASDAQ stocks with a minimum of two years' continuous return observations. Following Shumway (1997) and Shumway and Warther (1998), the data set is adjusted for the delisting bias. Monte Carlo realizations of bootstrap samples are independently generated to obtain 1,000 bootstrap pricing errors based on the CAPM for each sample stock. The empirical distribution of these bootstrap pricing errors is then used to construct the bootstrap confidence intervals. The 95% and 99% percentile confidence intervals used in this study are defined by the 2.5 and 97.5 percentiles and the 0.5 and 99.5 percentiles, respectively, of the distribution of the bootstrap pricing errors. If the confidence interval covers the hypothesized value of zero, then the null hypothesis will be accepted; if the interval does not cover zero, then the null hypothesis will be rejected.

		Bootstrap Distribution (%)			
Test Period	# of Sample Stocks	0.5	99.5	Total	Z-Statistic
Panel A: 1% Level of Significance					
		Rejection Rate (%)			
1948-1952	939	0.32	1.60	1.93	0.56
1953-1957	1010	0.89	2.48	3.37	1.94
1958-1962	1015	0.59	0.69	1.28	-1.23
1963-1967	1728	1.62	0.23	1.85	0.80
1968-1972	1981	1.21	1.77	2.98	3.35*
1973-1977	4155	2.09	0.29	2.38	2.44*
1978-1982	4119	2.06	0.53	2.59	2.45*
1983-1987	4496	1.07	1.56	2.63	4.50*
1988-1992	4911	1.53	1.28	2.81	4.39*
1993-1997	4414	1.29	1.06	2.35	3.57*
Mean		1.47	1.04	2.51	
		Bootstrap Distribution (%)			
		2.5	97.5	Total	
Panel A: 5% Level of Significance					
1948-1952	939	2.45	5.11	7.56	1.71
1953-1957	1010	3.76	5.25	9.01	2.35*
1958-1962	1015	4.33	2.66	6.99	0.08
1963-1967	1728	7.29	1.91	9.20	3.75*
1968-1972	1981	3.23	4.24	7.47	2.38*
1973-1977	4155	7.82	0.72	8.54	3.31*
1978-1982	4119	6.99	1.38	8.37	3.31*
1983-1987	4496	3.98	4.69	8.67	5.29*
1988-1992	4911	4.44	3.26	7.70	2.42*
1993-1997	4414	4.33	2.51	6.84	3.03*
Mean		5.20	2.83	8.03	

*Significant at the 5% level of significance

null hypothesis, we add artificial monthly returns to the return observations and study the bootstrap rejection rates under the alternatives. Figure 2.1 shows the bootstrap rejection rates of the GLPM and the CAPM when +1% to +5%, in increments of 1%, of artificial monthly returns are added to sample stock returns. It is clear that the bootstrap rejection rates increase under the two models as the level of artificial returns increases. It also appears that the bootstrap rejection rates of the two models are quite similar under the alternatives. Similar results can also be found in figure 2.2 when −5% to −1%, in increments of 1%, of artificial monthly returns are added to sample stock returns.

2.5.5 The Distribution of the Rejection Rates across Sizes

Focusing hypothesis testing on equation (2.2) and H_0: $PE_i = 0$, we show that, during the past 50 years from 1948 through 1997, no more than 3% of US stocks are mispriced. It would be interesting to see whether the GLPM can do an equally good job in explaining the pricing of small and big size stocks.

The data set used for examining the distribution of the rejection of the null hypothesis across sizes is similar to the one described in section 2.3. It comprises the set of all delisting-bias-adjusted NYSE, AMEX, and NASDAQ stocks from January 1948 to December 1997. Sample stocks in each five-year test period need to have at least two years' continuous returns. For each test period, based on NYSE size breakpoints, all sample stocks are sorted into one of the 10 portfolios, S1 to S10, arranged in order of increasing market capitalization at the portfolio formation month which is defined as the December immediately prior to the test period. In addition, to be qualified as sample stocks, they need to have continuous returns for the four

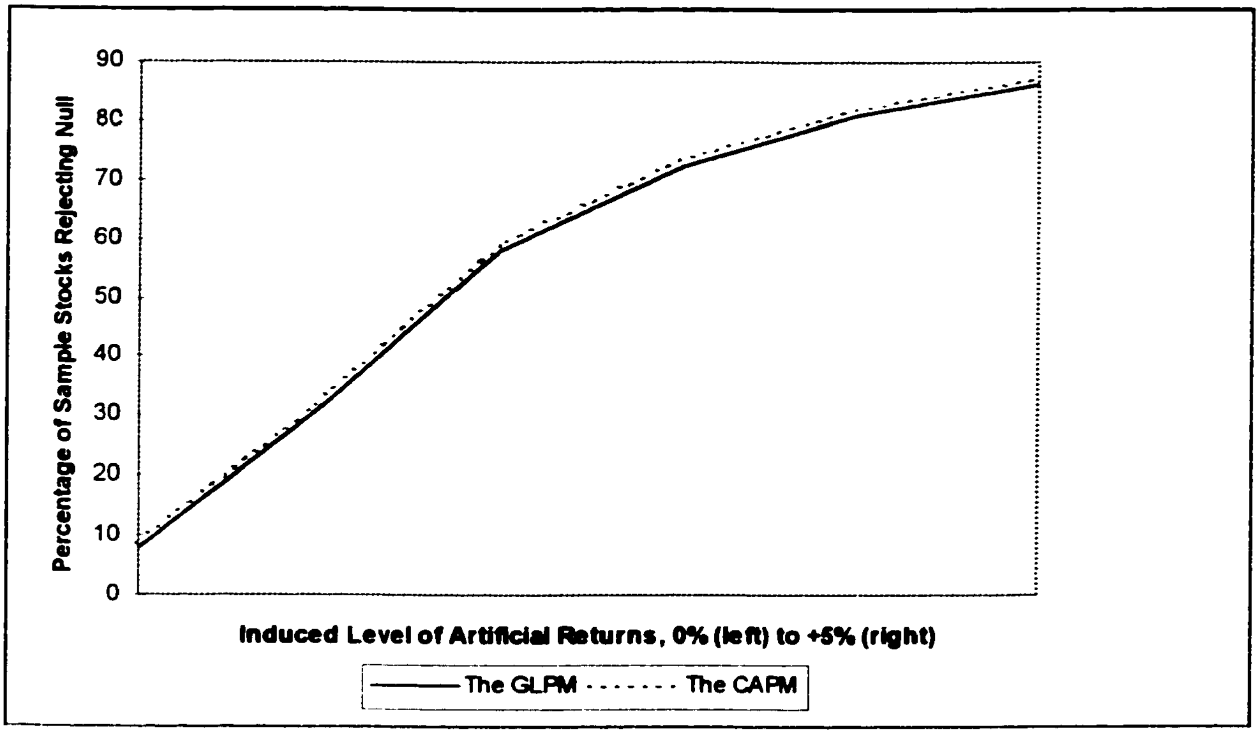

Figure 2.1 The bootstrap rejection rates based on the GLPM and the CAPM with positive artificial returns added. The percentage of sample stocks rejecting the null hypothesis of mispricing under the alternative hypothesis at various induced levels of positive artificial monthly returns (0% to +5% in increments of 1%) based on the GLPM and the CAPM. The bootstrap tests are one-tailed at the 5% level of significance.

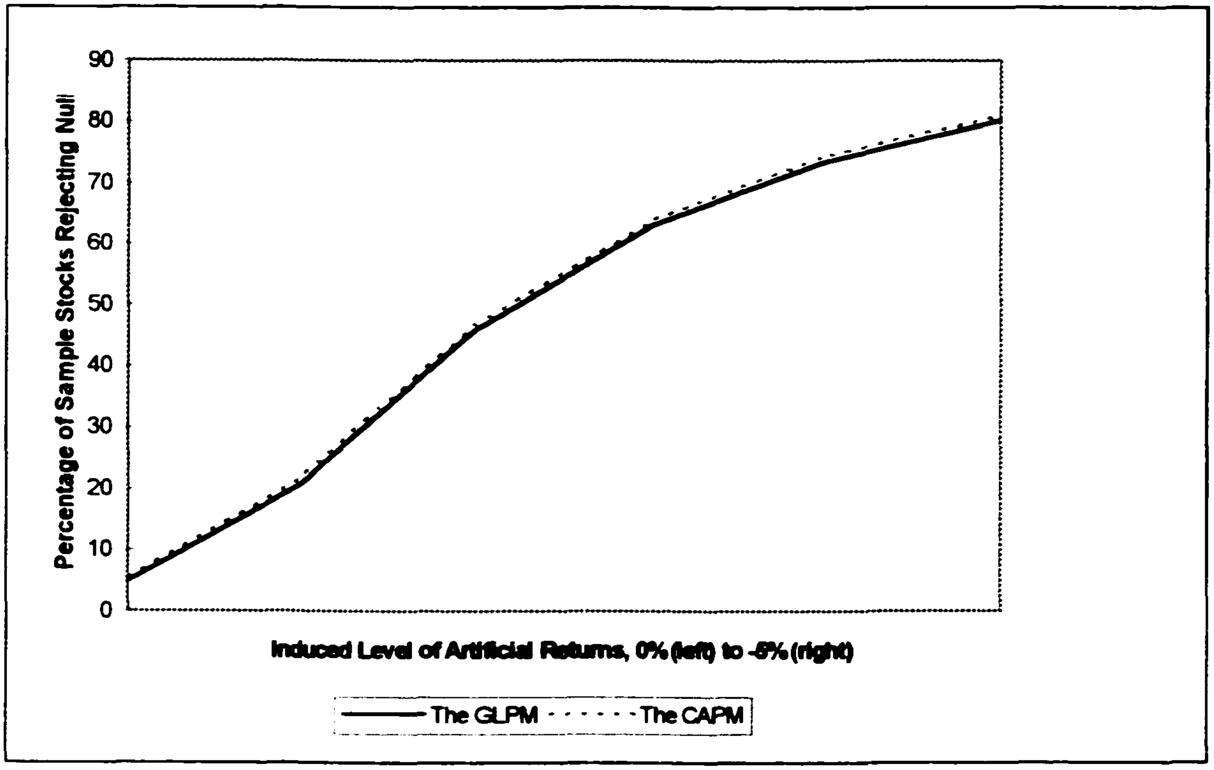

Figure 2.2 The bootstrap rejection rates based on the GLPM and the CAPM with negative artificial returns added. The percentage of sample stocks rejecting the null hypothesis of mispricing under the alternative hypothesis at various induced levels of negative artificial monthly returns (0% to –5% in increments of 1%) based on the GLPM and the CAPM. The bootstrap tests are one-tailed at the 5% level of significance.

years prior to the portfolio formation date to mitigate the complication from the IPO underperformance documented by Ritter (1991) and the others.[15]

Table 2.7 reports the distribution of the bootstrap rejection rates of the null hypothesis H_0: $PE_t = 0$ for the 10 size portfolios. During the 1948-1997 sample period, the density function of the bootstrap rejection rates across sizes is decreasing in size, but not monotonically. At the 1% level of significance, the rejection rates for S1 to S10 portfolios during the 1948-1997 sample period are 2.74%, 2.37%, 2.67%, 1.65%, 1.88%, 1.94%, 1.72%, 1.14%, 0.88%, and 1.09%, respectively. At the 5% level of significance, the rejection rates across sizes are 7.91%, 8.33%, 7.54%, 6.60%, 5.96%, 6.71%, 5.11%, 5.52%, 3.81%, and 4.64%, from the smallest size decile to the largest one. Compared with the mean rejection rates of 2.03% and 7.04% in table 2.5 for the 1% and 5% levels of significance, respectively, the rejection rates for S4 to S10 portfolios appear to be moderate. On the other hand, the rejection rates for S1 to S3 portfolios are above the mean rejection rates of 2.03% and 7.04% for the 1% and 5% levels of significance, respectively. These patterns are consistent with the documented size effect, in the sense that mispricing occurs among small stocks. We also document that small stocks tend to generate positive excess returns so that the rejection of the null hypothesis tends to occur at the downside of the bootstrap distributions.

However, given the mean rejection rates of 2.03% and 7.04% in table 2.5, even the rejection rates of 2.74% and 7.91% in S1 portfolio do not appear too far

[15] The conclusions are invariant to whether the sample stocks are required to have continuous returns for the four years prior to the portfolio formation month.

Table 2.7
The Distribution of the Bootstrap Rejection Rates across Sizes

This table presents the rejection rates of the null hypothesis H_0: $PE_i = 0$ with bootstrap confidence intervals for the 10 size portfolio during the 1948-1997 sample period. The data comprises the set of all NYSE, AMEX, and NASDAQ stocks with a minimum of two years' continuous return observations. Following Shumway (1997) and Shumway and Warther (1998), the data set is adjusted for the delisting bias. For each five-year test period, based on NYSE size breakpoints, all sample stocks are sorted into one of the 10 portfolios, S1 to S10, arranged in order of increasing market capitalization at the portfolio formation month. In addition, to be qualified as sample stocks, they need to have continuous returns for the four years prior to the portfolio formation date. Monte Carlo realizations of bootstrap samples are independently generated to obtain 1,000 bootstrap pricing errors for each sample stock. The empirical distribution of these bootstrap pricing errors is then used to construct the bootstrap confidence intervals. The 95% and 99% percentile confidence intervals used in this study are defined by the 2.5 and 97.5 percentiles and the 0.5 and 99.5 percentiles, respectively, of the distribution of the bootstrap pricing errors.

			Two-Tailed Theoretical Significance Level					
		# of	1%			5%		
Size Decile		Sample	Bootstrap Distribution (%)					
		Stocks	0.5	99.5	Total	2.5	97.5	Total
			Rejection Rate (%)					
Smallest	1	5098	1.53	1.21	2.74	5.28	2.63	7.91
	2	1813	1.43	0.94	2.37	5.63	2.70	8.33
	3	1538	1.63	1.04	2.67	4.49	3.05	7.54
	4	1396	1.07	0.58	1.65	4.16	2.44	6.60
	5	1270	0.94	0.94	1.88	3.83	2.13	5.96
	6	1178	0.50	1.44	1.94	3.06	3.65	6.71
	7	1095	0.91	0.82	1.72	3.01	2.10	5.11
	8	1052	0.29	0.85	1.14	2.95	2.57	5.52
	9	1024	0.39	0.49	0.88	1.85	1.96	3.81
Largest	10	1014	0.30	0.79	1.09	2.17	2.47	4.64
Mean			0.90	0.91	1.81	3.65	2.57	6.22

away from the 1% and 5% levels of significance, respectively, under the null. In addition, in comparison with the downside 0.5% and 2.5% levels of significance, the downside rejection rates of 1.53% and 5.28% in S1 portfolios suggest that it will be difficult to identify mispricing opportunities since, on average, at most three out of 100 stocks in the smallest size decile are mispriced. Overall, these results suggest that the mispricing in small size portfolios is not severe.

2.5.6 Biases of the LS Estimates

Instead of focusing hypothesis testing on equation (2.2) and H_0: $PE_t = 0$, one may attempt to examine the GLPM by testing a linear functional form, such as equation (2.4) or (2.5). To illustrate this linear testing framework, we assume that:

$$\hat{E}(MRG_t) = E(MRG_t) + \xi_t$$

$$\hat{E}(MRL_t) = E(MRL_t) + \delta_t$$

where estimation errors ξ_t and δ_t are assumed to be normally distributed with zero means and uncorrelated with the true values, $E(MRG_t)$ and $E(MRL_t)$. These error assumptions are then substituted into the GLPM to obtain the following identity:

$$\hat{E}(MRG_t) = \pi \hat{E}(MRL_t) + \xi_t - \pi \delta_t$$

This equation then can be rewritten as a regression model:

$$\hat{E}(MRG_t) = a + b \hat{E}(MRL_t) + e_t \qquad (2.11)$$

where $a = 0$, $b = \pi$, $e_t = \xi_t - \pi\,\delta_t$. In practice, the estimated coefficients $\hat{a}$ and $\hat{b}$ are compared to zero and an estimate of π estimated from the same estimation period, respectively. That is, this regression model can be estimated for each test period based on the two-stage cross-sectional regression of Fama and MacBeth (1973). However, the regression model is subject to EIV (errors-in-variable) problems in which the second stage regressors are estimated with errors.

To demonstrate the extent of EIV biases, the results of the LS regression analysis on equation (2.11) are reported in table 2.8. As expected, the estimates $\hat{b}$ in the 10 five-year test periods are all biased downward when they are compared with the sample estimates $\hat{\pi} = \hat{E}(G_m)/\hat{E}(L_m)$ from the same estimation period. For example, during the 1993-1997 test period, $\hat{b}$ is 0.2899 with a standard error of 0.0297, which is biased downward by approximately 88.34%, as measured by the percentage sample estimated bias, $(\hat{b} - \hat{\pi}) / \hat{\pi}$.

To see whether the downward biases in $\hat{b}$ are due to EIV problems, the percentage implied asymptotic bias is defined as:

$$(\hat{b} - \pi) / \pi = -\frac{\sum_N \sigma(\delta)^2}{N\sigma(E(\hat{ARL}_i))^2} \tag{2.12}$$

which is based on the asymptotic LS estimator of π (Judge, Griffiths, Hill, and Lee, 1980, p. 514):

$$plim\ \hat{b} = \pi - \pi \frac{\sum_N \sigma(\delta)^2}{N\sigma(E(\hat{ARL}_i))^2} \tag{2.13}$$

Table 2.8
Biased Results of LS Regressions, 1948-1997

This table presents the results based on the following regression specification:

$$\hat{E}(MRG_i) = a + b\ \hat{E}(MRL_i) + e_i$$

where $\hat{E}(MRG_i) = E(MRG_i) + \xi_i$; and $\hat{E}(MRL_i) = E(MRL_i) + \delta_i$. Let $\pi = \hat{E}(G_m) / \hat{E}(L_m)$. The percentage sample estimated bias is defined as $(\hat{b} - \pi) / \pi$. The percentage implied asymptotic bias is defined as:

$$(\hat{b} - \pi) / \pi = -\frac{\sum_{N} \sigma(\delta)^2}{N\sigma(E(MRL_i))^2}$$

The data comprises the set of all NYSE, AMEX, and NASDAQ stocks with a minimum of two years' continuous return observations. Following Shumway (1997) and Shumway and Warther (1998), the data set is adjusted for the delisting bias.

Test Period	a (Std. Err.)	$\hat{b}$ (Std. Err.)	π	Percentage Estimated Bias	Percentage Implied Asymptotic Bias	R^2(%)
1948-1952	0.0178	0.5127	2.2700	-77.41%	-62.51%	13.81
	(0.0006)	(0.0418)				
1953-1957	0.0136	0.3193	1.8922	-83.13%	-72.62%	7.90
	(0.0004)	(0.0343)				
1958-1962	0.0173	0.4150	1.8545	-77.62%	-64.99%	9.95
	(0.0006)	(0.0392)				
1963-1967	0.0278	0.2154	2.0130	-89.30%	-104.11%	0.89
	(0.0008)	(0.0547)				
1968-1972	0.0140	0.3335	1.1220	-70.28%	-66.50%	6.14
	(0.0008)	(0.0293)				
1973-1977	0.0219	0.1091	0.8148	-86.61%	-73.13%	1.12
	(0.0005)	(0.0159)				
1978-1982	0.0259	0.2832	1.2697	-77.70%	-91.04%	2.16
	(0.0006)	(0.0297)				
1983-1987	0.0186	0.0641	1.4314	-95.52%	-104.29%	0.20
	(0.0007)	(0.0212)				
1988-1992	0.0177	0.0772	1.6964	-95.45%	-96.99%	0.32
	(0.0005)	(0.0194)				
1993-1997	0.0168	0.2899	2.4868	-88.34%	-88.44%	2.12
	(0.0004)	(0.0297)				
Mean	0.0191	0.2619	1.6851	-84.14%	-82.46%	4.46
	(0.0006)	(0.0315)				

If the biases in b are largely due to EIV problems and the sample estimates $\hat{\pi}$ are a good proxy for the parameter π, the percentage implied asymptotic biases, $(\hat{b} - \pi)$ / π, should be close to the percentage sample estimated biases, $(\hat{b} - \hat{\pi}) / \hat{\pi}$. Using the sample variances for $\sigma(\delta_i)^2$ and $\sigma(\hat{E}(MRL_i))^2$, the percentage implied asymptotic biases are reported in table 2.8. During the 1948-1997 sample period, the percentage implied asymptotic biases are all quite close to and correlated with the percentage sample estimated biases. These results indicate that EIV problems are severe when tests are based on a two-stage cross-sectional regression.

2.6 Conclusions

This study tests the GLPM whose functional form is different from the traditional beta-coefficient functional form. The GLPM incorporates loss aversion, in which the risk-return tradeoffs are based on expected market-related losses and expected market-related gains. The testing framework used in this study also imposes a higher testing standard and places no restrictions on the distribution of the statistic with the use of the bootstrap. The results show that the mean rejection rates of the GLPM are 2.03% and 7.04% for the 1% and 5% levels of significance, respectively, during the 1948-1997 sample period. The highest rejection rates in the 10 five-year test periods are 3.07% and 8.18% for the 1% and 5% levels of significance, respectively. Given the allowable level of the 1% and 5% levels of significance and the randomness in the level of the type-I errors, no more than 3% of sample stocks are mispriced. Moreover, the mispricing in small size portfolios is not severe. These results indicate that the GLPM appears to do a good job of explaining stock returns.

Based on the CAPM, researchers have identified a number of anomalies in asset prices. Empirical anomalies can arise from several sources, such as market frictions, false inferences of the risk-return relationship, inappropriate risk measures, and mispricing by market participants. The results of the bootstrap tests in this study suggest that there are not many mispriced stocks according to the GLPM. For future research, it should be interesting to apply this testing framework, based on the GLPM, to re-examine these anomalies.

CHAPTER 3

LONG-TERM PERFORMANCE EVALUATION WITH A NORMATIVE ASSET PRICING MODEL

3.1 Introduction

Recent empirical studies document a number of anomalies in long-term common stock performance.[16] While the empirical evidence appears quite strong, long-term performance studies that use *ad hoc* models suffer from theoretical and statistical difficulties, which limits their usefulness as tests of the efficient market hypothesis. This study provides a new long-term testing framework that mitigates these theoretical and statistical problems.

To study long-term performance, researchers need an asset pricing model to generate "normal" (expected) returns so that "abnormal" returns can be measured. Because of the prominent results of Fama and French (1992, 1993, 1996), many recent long-term performance studies use the Fama-French three factor model to obtain expected returns. However, the Fama-French three factor model is a positive (empirically based) model. The model empirically identifies the mimicking portfolios of size and book-to-market ratio (*BE/ME*) as risk factors. According to Loughran and Ritter (2000), "if a positive (empirically based) model is used, one is not testing market efficiency; instead, one is merely testing whether any patterns that exist are being captured by other known patterns." They argue that tests of market efficiency require a normative (equilibrium) asset pricing model.

[16] These anomalies include market overreaction (De Bondt and Thaler, 1985), the long-term underperformance of initial public offerings (Ritter, 1991), market underreaction (Jegadeesh and Titman, 1993), and the long-term underperformance of seasoned equity offerings (Loughran and Ritter (1995), Spiess and Affleck-Graves (1995)).

In addition, Fama (1998) and Fama and French (1996) demonstrate that long-term performance evaluation is sensitive to the assumed model for expected returns and argue that many long-term anomalies are artifacts of bad-model problems. Specifically, Fama (1998) argues that a reasonable change of models often causes an anomaly to disappear. Furthermore, as Fama (1998) indicates, even the Fama-French three-factor model does not provide a full explanation of returns in the size and *BE/ME* dimensions that are supposed to be well captured by construction. Therefore, it is of interest to see how long-term anomalies behave under another asset pricing model. To address these theoretical issues, in this study we use a new normative (equilibrium) model to assess long-term performance.

This new normative model incorporates loss aversion into equilibrium asset pricing. Empirical evidence on risk aversion indicates that the variance of portfolio returns does not fully characterize economic agents' perception of and behavior toward risk. A number of studies of risky choice involving monetary outcomes have documented that economic agents are averse to losses. For example, Kahneman and Tversky (1979) and Tversky and Kahneman (1986, 1991, 1992) find that economic agents are much more sensitive to losses than to gains. This loss-averse view is also advocated by Fishburn (1977) and Benartzi and Thaler (1995).

The property of loss aversion is specified by a gain-loss utility function in this new normative (equilibrium) model. Under this utility function restriction, in which investors are assumed to be more averse to losses when they expect to lose more, Lin (1999a) derives the Gain-Loss Pricing Model (GLPM). The GLPM is more robust than the Capital Asset Pricing Model (CAPM) of Sharpe (1964), Lintner (1965), and

Mossin (1966), and is valid for all financial assets at the Pareto-optimal allocation of risk. Because of its theoretical properties and normative nature, we use the GLPM for the measurement of pricing errors.

It is well known that the traditional event-time methods for long-term performance evaluation are troubled by the skewness of abnormal return distributions. Barber and Lyon (1997), Kothari and Warner (1997), Fama (1998), and Lyon, Barber, and Tsai (1999) show that commonly used return metrics to test for long-term abnormal stock returns may lead to misspecified test statistics due to the skewness of abnormal return distributions. These inference problems are particularly severe when monthly returns are compounded to obtain abnormal return measures.

To resolve these statistical problems, this study uses the nonparametric bootstrap method that incorporates the distributional information intrinsic within the sample into estimation and inference. The nonparametric bootstrap is free from distributional assumptions; it only requires a random sample and a proposed statistic. Thus, the statistical difficulties due to skewness of abnormal return distributions are mitigated. Furthermore, the long-term performance evaluation framework used in this study is based on a return metric obtained by resampling return observations and by analyzing the whole return distribution without compounding performance measures. Consequently, our test is free from the aggregation problems in many long-term performance tests.

The specification of our long-term performance evaluation framework is evaluated using samples of randomly selected NYSE/AMEX/NASDAQ stocks and

simulated random event dates. Our simulation results show that the long-term performance evaluation framework based on the GLPM is well-specified. When artificial abnormal returns are added to real return observations, we find that our evaluation framework exhibits adequate power to reject the null hypothesis of no long-term abnormal returns. We further illustrate the use of this long-term performance evaluation framework by examining the long-term returns of initial public offerings (IPOs). We find that, consistent with the documented underperformance (Ritter (1991), Loughran and Ritter (1995), Spiess and Affleck-Graves (1995), and Loughran and Ritter (2000)), IPOs tend to be negatively mispriced in their post-event windows according to the GLPM.

The remainder of this essay is organized as follows. Section 3.2 introduces the GLPM and characterizes the normative (equilibrium) nature of the GLPM. Section 3.3 explains the long-term evaluation framework, the bootstrap method, and simulation methods. Section 3.4 provides simulation results. Section 3.5 illustrates the use of this approach by examining the long-term returns of IPOs. Section 3.6 concludes this essay.

3.2 The Gain-Loss Pricing Model (GLPM)

The GLPM postulates that any asset *i* must be priced such that:

$$\frac{E(MRG_i)}{E(MRL_i)} = \frac{E(G_m)}{E(L_m)} \tag{3.1}$$

where

$$MRG_i = (R_i - R_f) \text{ if } R_m > R_f \text{ and } MRG_i = 0 \text{ otherwise;}$$

$$MRL_i = -(R_i - R_f) \text{ if } R_m < R_f, \text{ and } MRL_i = 0 \text{ otherwise;}$$

$$G_m = (R_m - R_f) \text{ if } R_m > R_f, G_m = 0 \text{ otherwise; and}$$

$$L_m = -(R_m - R_f) \text{ if } R_m < R_f, L_m = 0 \text{ otherwise.}$$

As a normative (equilibrium) model, the GLPM says that, in market equilibrium, the ratio of expected market-related gain to expected market-related loss of asset i, $E(MRG_i)/E(MRL_i)$, is equal to the ratio of expected gain to expected loss of the market portfolio, $E(G_m)/E(L_m)$. Since the gain-loss ratio of the market portfolio, $\pi \equiv E(G_m)/E(L_m)$, is invariant to all assets, the GLPM predicts that the higher the expected market-related loss of an asset, the higher the expected market-related gain investors require to compensate for the loss. Thus, under the GLPM, the gain-loss ratio of the market portfolio defines the gain-loss tradeoff for individual assets.

3.3 The Long-Term Performance Evaluation Framework

In this section, our long-term performance evaluation framework is introduced. The selection of critical values based on the bootstrap percentiles and the hypothesis testing for each individual security are first outlined. Then a test statistic for a sample of finitely many event securities is proposed. We close this section with a description of the simulation method used to evaluate the specification and the power of the long-term testing framework.

3.3.1 The Test Statistic for Each Security

The long-term performance evaluation framework proposed in this study comprises two steps. The first step is to test whether the long-term performance of an event security and a matching pseudo security is significantly deviated from the GLPM risk-return tradeoff in the post-event window. This testing procedure is

repeated for each event and pseudo security to obtain two rejection rates of the GLPM risk-return tradeoff for the event and pseudo samples, one rejection rate for each sample. The second step of the long-term performance evaluation framework is then to test the equality of the two rejection rates between the event sample and the pseudo sample.

According to the GLPM, the pricing error for security i, $PE_i = E(MRG_i)$ $E(L_m) - E(MRL_i)\, E(G_m)$, should be zero if the pricing of the security is consistent with the GLPM. Hence, if we are interested in knowing whether a particular event security exhibits long-term positive or negative abnormal performance, we can conduct a direct test by testing the null hypothesis H_0: $PE_i \leq 0$ or H_0: $PE_i > 0$, respectively, for the event security. However, PE_i is a function of the unknown distributions of MRG_i, L_m, MRL_i, and G_m. For long-term performance evaluation, we are also only interested in the pricing of an event security in the post-event window. Correspondingly, we define the pricing error statistic, $\hat{PE}_i$, as:

$$\hat{PE}_i = \hat{E}(MRG_i)\,\hat{E}(L_m) - \hat{E}(MRL_i)\,\hat{E}(G_m) \tag{3.2}$$

and the statistic $\hat{PE}_i$ depends on the return observations in i's post-event window. It is also understood that post-event window is individual-specific; that is, different event securities may have different post-event windows.

3.3.2 The Bootstrap Distribution and Hypothesis Testing on Each Security

A primary question in long-term event studies is whether one particular event has a long-term economic impact on the value of a firm. If the economic impact soon dissipates or even does not exist, the empirical risk-return tradeoff based

on the post-event return *distributions* should be consistent with a well-specified asset pricing model because at most only a few return observations deviate from the benchmark. That is, given a size n of event sample and the r level of significance, we should observe that there are approximate $n \times (1 - r)$ event stocks whose pricing in post-event windows is consistent with the well-specified asset pricing model. In contrast, if the economic impact persists and many return realizations systematically deviate from the benchmark, we should observe that the empirical risk-return tradeoff based on the post-event return *distributions* will be inconsistent with the well-specified asset pricing model. That is, the number of post-event mispricing will be significantly higher than $n \times (1 - r)$.

In this study, we focus on sample pricing error $\hat{PE}_i$ which is defined by the post-event return *distributions* and use the nonparametric bootstrap method to estimate the distributional properties of $\hat{PE}_i$. Based on Monte Carlo resampling of return observations, this method simulates the bootstrap pricing errors, $\hat{PE}_i*$. To test whether a security has long-term positive or negative abnormal returns in its post-event window, our testing framework uses the empirical bootstrap distribution of $\hat{PE}_i*$ to find the critical values. Let H be the empirical bootstrap distribution of $\hat{PE}_i*$. The r% critical value for testing whether a security has long-term positive abnormal returns is defined by the r percentile of $\hat{H}$. For example, with a 1,000 times of Monte Carlo simulation, the 5% critical value is the 50th observation of the ranked $\hat{PE}_i*$. Similarly, at the 1% level of significance, the critical value is the 10th observation of the ranked $\hat{PE}_i*$. Once the critical values are obtained, statistical inferences can be made in the usual way. If the r critical value is greater than the

hypothesized value of zero, the null hypothesis H_0: $PE_i \leq 0$ is accepted for event security i; if the critical value is less than zero, the alternative hypothesis H_0: $PE_i > 0$ is accepted for event security i. Similarly, to test whether a security has long-term negative abnormal returns, the r% critical value is defined by the $1 - r$ percentile of H. Then, by repeating this testing procedure for each event sample security, we obtain the rejection rate of the tests for long-term positive or negative abnormal performance for the event sample.

We use the empirical critical values based on the nonparametric bootstrap because the nonparametric bootstrap is free from distributional assumptions; it only requires a random sample and a proposed statistic of PE_i. This property is particularly appealing for this study since the distribution F is unknown and may not belong to a class of analytically tractable distributions. The strength of the bootstrap method is that, even if the distribution F is unknown, this method leads to a consistent estimator of PE_i (Efron, 1979). Furthermore, the traditional event-time methods for long-term performance evaluation are hampered by the skewness of abnormal return distributions. By incorporating the distributional information intrinsic to the sample, the statistical inferences based on the bootstrap method used in this study are likely to be more robust.

3.3.3 The Test Statistic for a Pair of n Event Securities and n Pseudo Securities

After acquiring the rejection rates of the null hypotheses H_0: $PE_i \leq 0$ and H_0: $PE_i \geq 0$ for the event sample of size n, we construct a pseudo sample of size n from non-event securities and to obtain the rejection rates of the null hypotheses H_0: $PE_i \leq 0$ and H_0: $PE_i \geq 0$ for the pseudo sample. The construction of the pseudo

sample matches the sampling timing of the event sample. That is, for each event security, a pseudo security is randomly selected from non-event securities such that the pseudo security has return observations covering the beginning of the event security's post-event window. We then include the available return observations of the pseudo security in the post-event window in the pseudo sample. The reason that the event and pseudo samples are matched in the beginning of the sampling period is that the sampling of the event securities is not independent. For example, the hot market hypothesis of IPOs suggests that IPOs cluster in calendar time. It is widely known that certain events typically cluster in calendar time and this clustering affects statistical inferences (Brown and Warner (1980, 1985)). To the extent that calendar-time clustering exists in the event sample, it is also present in the pseudo sample, and thus is, to some degree, controlled for in our tests. Also, this matching procedure avoids the survivorship bias in the sense that the survivorship of event and pseudo securities are independent of each other. Therefore, with the same probability, event securities may have a shorter or longer return series than the corresponding pseudo securities.

When choosing a pseudo security for an event security, we exclude the other event securities whose post-event windows are within the event security's post-event window from consideration. Based on Loughran and Ritter's (2000) argument, the least powerful test is to select a pseudo security only from the event securities. Therefore, to enhance the power of the test, we make sure that the pseudo sample is not contaminated by the presence of event securities.

The second step of the long-term performance evaluation framework is to test the equality of the two rejection rates between the event sample and the pseudo sample. Under the null hypothesis that there is no long-term abnormal returns associated with the event, the two rejection rates should be equal. In contrast, if event securities tend to generate positive (negative) abnormal long-term returns, then the event sample should have a higher rejection rate with the null hypothesis H_0: $PE_t \leq 0$ (H_0: $PE_t \geq 0$) than the pseudo sample.

We use the Z-test for the equality between two proportions (binomial distribution). This test is used to investigate the null hypothesis of equality between two population mispricing proportions, based on two sample mispricing rates, κ_{event} and κ_{pseudo}. The test statistic is:

$$Z = \frac{\kappa_{event} - \kappa_{pseudo}}{\{P(1-P)\left[\frac{2}{n}\right]\}^{1/2}} \tag{3.3}$$

where

$$P = \frac{\kappa_{event} + \kappa_{pseudo}}{2}$$

Under the null hypothesis H_0: $PE_t \leq 0$, κ_{event} and κ_{pseudo} are the rejection rates of H_0: $PE_t \leq 0$ based on the r percentile of $\dot{H}$ for event and pseudo sample, respectively. Under the null hypothesis H_0: $PE_t \geq 0$, κ_{event} and κ_{pseudo} are the rejection rates of H_0: $PE_t \geq 0$ based on the $1 - r$ percentile of $\dot{H}$ for event and pseudo sample, respectively. Under both null hypotheses, Z is approximately distributed as a unit normal distribution when the sample size is sufficient large, e.g., $n \geq 30$. Given that

the sample sizes in most event studies are much higher than this number, the use of the Z-test should be appropriate in practice.

3.3.4 Simulation Method

To test the specification of our long-term performance evaluation framework, 500 random event samples of 200 sample securities each are drawn. We randomly select event securities with replacement from the population of all NYSE, AMEX, and NASDAQ ordinary common stocks in the 1997 Center for Research in Security Prices (CRSP) monthly files. Our analysis covers the period July 1973 through December 1997. We chose this period because we would like to compare our results with the power tests documented in Lyon, Barber, and Tsai (1999) whose sample period starts in July 1973.

Once an event stock is selected, we adjust its return observations for the delisting bias to mitigate survivorship bias. Shumway (1997) finds that most of the missing delisting returns in the CRSP tapes are associated with negative events and suggests a –30% delisting monthly return for NYSE and AMEX stocks. Similarly, Shumway and Warther (1998) suggest a corrected delisting return of –55% for NASDAQ stocks. Following Shumway (1997) and Shumway and Warther (1998), we classify delisting codes 500 and 505 through 588 as negative-performance-related and adjust the data set for the missing delisting monthly returns with –30% for NYSE and AMEX stocks and with –55% for NASDAQ stocks.

Each time an event stock is selected, we randomly generate a random event month between July 1973 and December 1992, December 1994, or December 1996, depending on whether performance is evaluated over a five-, three-, or one-year

period following the event month. To mitigate the effect of survivorship bias, if the event stock does not have full observations, the bootstrap test is conducted for as many months as the delisting-bias-adjusted data are available. If the event stock does not have return observations for the event month, we randomly select another event stock with another event month until the event stock has return observations for the event month. By doing so, we are more likely to sample stocks with a longer period of return observations. As Barber and Lyon (1997) argue, this is sensible since most long-term event studies analyze events that are proportional to the history of a firm.

For each event stock, a pseudo stock is randomly chosen with the delisting bias adjusted from the non-event stocks in the 1997 CRSP population such that the available return observations of the pseudo stock cover the beginning of the post-event window of the event stock. That is, for the pseudo stock, we do not generate another event month; we include the available return observations of the pseudo security in its post-event window in the pseudo sample. Therefore, the probability that the event stock may have a higher or lower number of available return observations than the pseudo stock is the same. By repeating this matching procedure, we obtain 500 pseudo samples of 200 pseudo stocks each. Once the event and pseudo samples are constructed, we apply the bootstrap method to each event and pseudo stock.

The next step of the long-term testing framework is to use the Z-test in equation (3.3) with the α level of significance for each pair of the 500 event and pseudo samples. If the test is well-specified, 500α tests should reject the null hypothesis of equality between two population mispricing proportions because of the

Type-I error. We test the specification of the long-term testing framework at the 1% and 5% levels of significance.

3.4 Simulation Results

In this section, we present the simulation results in 500 random pairs of event and pseudo samples. The specification and power of the long term performance evaluation framework are the focus of this study.

3.4.1 Specification

The baseline results are based on 500 random pairs of event and pseudo samples with 200 random events in each pair. To assess the specification of the long-term performance evaluation framework based on the GLPM, we first test whether the GLPM describes the risk-return tradeoff in the post-event window for each event and pseudo stock. Table 3.1 reports the average rejection rates of the null hypothesis of no long-term positive abnormal returns, i.e., H_0: $PE_t \leq 0$, and that of no long-term negative abnormal returns, i.e., H_0: $PE_t \geq 0$, for the 100,000 (500 × 200) event stocks and the corresponding 100,000 pseudo stocks at one-, three-, and five-year horizons. At the 1% level of significance, the empirical rejection rates range from 1.28% to 2.46%. The empirical rejection rates range from 5.57% to 8.77% for the 5% level of significance. The test results show that the rejection rates of the bootstrap test are not far from the theoretical level of significance. The rejection rates for H_0: $PE_t \leq 0$ and H_0: $PE_t \geq 0$ seem, in general, not far from symmetry; the only exception is that when the 5% level of significance is used at the five-year horizon we observe a higher frequency of positive abnormal performance. It also appears that our long-term performance evaluation framework is not

Table 3.1
Rejection Rates of the Null Hypotheses of No Long-Term Positive and Negative Abnormal Returns with Bootstrap Critical Values on Each Event and Pseudo Stock for the 500 Random Pairs of Event and Pseudo Samples of 200 Stocks Each

This table presents the rejection rates of the null hypotheses of no long-term positive abnormal returns , i.e., H_0: $PE_t \leq 0$, and of no long-term negative abnormal returns, i.e., H_0: $PE_t \geq 0$, with bootstrap critical values on each random event stock and pseudo stock. We first randomly select event stocks with replacement from the population of all NYSE, AMEX, and NASDAQ ordinary common stocks in the 1997 CRSP monthly files. Once an event stock is selected, following Shumway (1997) and Shumway and Warther (1998), we adjust its return observations for the delisting bias to mitigate survivorship bias. For each event stock, a random event month between July 1973 and December 1992, December 1994, or December 1996 is generated, depending on whether performance is evaluated over a five-, three-, or one-year period following the event month. This procedure is repeated to obtain 500 random event samples of 200 event stocks each. Then, for each event stock, a pseudo stock is randomly chosen with the delisting bias adjusted from the other stocks in the 1997 CRSP population such that the available return observations of the pseudo stock cover the beginning of the post-event window of the event stock. This procedure is repeated to obtain 500 random pseudo samples of 200 pseudo stocks each. Monte Carlo realizations of bootstrap samples are then independently generated to obtain 1,000 bootstrap pricing errors for each event and pseudo stock. The empirical distribution of these bootstrap pricing errors is used to construct the bootstrap critical values. The numbers presented are the rejection rates of the null hypotheses of no long-term positive abnormal returns , i.e., H_0: $PE_t \leq 0$, and of no long-term negative abnormal returns, i.e., H_0: $PE_t \geq 0$, on each random event stock and pseudo stock for the 500 pairs of event and pseudo samples.

	Theoretical Significance Level			
	1%		5%	
	The Null Hypothesis			
	H_0: $PE_t \leq 0$	H_0: $PE_t \geq 0$	H_0: $PE_t \leq 0$	H_0: $PE_t \geq 0$
Panel A. One-Year Horizon				
Event Sample	1.30	2.08	5.57	6.78
Pseudo Sample	1.28	2.11	5.64	6.77
Panel B. Three-Year Horizon				
Event Sample	2.26	2.17	8.08	6.33
Pseudo Sample	2.24	2.14	8.06	6.28
Panel C. Five-Year Horizon				
Event Sample	2.46	2.25	8.77	5.92
Pseudo Sample	2.41	2.23	8.71	5.93

particularly troubled by the skewness and aggregation problems since there is no monotonic relationship between empirical bootstrap rejection rate and the length of time horizon.

Under the null hypothesis, the specification of the long-term performance evaluation framework based on the GLPM is assessed by the Z-tests of equalities between the bootstrap rejection rates of H_0: $PE_t \leq 0$ and H_0: $PE_t \geq 0$ for the 500 random pairs of event and pseudo samples. The test results are reported in table 3.2. At the 1% level of significance for the Z-tests, the rejection rates of the equalities between the bootstrap rejection rates of H_0: $PE_t \leq 0$ and H_0: $PE_t \geq 0$ range from 0.2% to 2.0%. At the 5% level of significance for the Z-tests, the rejection rates of the equalities range from 4.2% to 7.2%.

If the rejection rates of the equalities between the bootstrap rejection rates is viewed as a random variable, the Z-test rejection rates will be close to, but not necessarily equal to, the level of significance for the Z-tests when the number of the pairs of event and pseudo samples is a fairly large finite number. According to Brown and Warner (1980), suppose that the testing outcomes for each pair of the N pairs of event and pseudo samples are independent, then at the a level of significance for the Z-tests, the rejection rate of the N Z-tests for such a Bernoulli process has a mean of a and a standard deviation of:

$$\sigma_{Bernoulli, (a, 1-a)} = \sqrt{\frac{a \times (1-a)}{N}} \tag{3.4}$$

Table 3.2
Test Results of the Null Hypothesis of the Equalities between the Rejection Rates of H_0: $PE_i \leq 0$ and H_0: $PE_i \geq 0$ for the 500 Random Pairs of Event and Pseudo Samples

This table presents the rejection rates of the null hypothesis of the equalities between the rejection rates of H_0: $PE_i \leq 0$ and H_0: $PE_i \geq 0$ for the 500 random pairs of event and pseudo samples during the sample period of July 1973 through December 1997. We first construct 500 random pairs of event and pseudo samples from the population of all NYSE, AMEX, and NASDAQ ordinary common stocks in the 1997 CRSP monthly files at one-, three-, and five-year horizons. Following Shumway (1997) and Shumway and Warther (1998), we adjust return observations for the delisting bias to mitigate survivorship bias. Monte Carlo realizations of bootstrap samples are then independently generated to obtain 1,000 bootstrap pricing errors for each event and pseudo stock. By the bootstrap tests, we obtain the rejection rates of the null hypothesis H_0: $PE_i \leq 0$ and H_0: $PE_i \geq 0$ for each event and pseudo sample. With the 500 pairs of rejection rates for H_0: $PE_i \leq 0$ and for H_0: $PE_i \geq 0$, we then apply the one-tailed Z-test for the equality between rejection rates to each pair of observations. The number reported are the rejection rates of the 500 Z-tests.

	Theoretical Significance Level			
	1%		5%	
	The Null Hypothesis			
	H_0: $PE_i \leq 0$	H_0: $PE_i \geq 0$	H_0: $PE_i \leq 0$	H_0: $PE_i \geq 0$
Panel A. Z-Test with 1% level of Significance				
One-Year Horizon	0.8	0.2	0.8	0.8
Three-Year Horizon	2.0*	0.6	1.2	0.8
Five-Year Horizon	0.8	1.4	1.6	0.8
Panel B. Z-Test with 5% level of Significance				
One-Year Horizon	5.4	6.8	5.6	5.6
Three-Year Horizon	6.0	5.4	5.0	4.6
Five-Year Horizon	5.0	7.2*	5.6	4.2

*Significant at the 5% level, two-tailed (Bernoulli process)

That is, we should expect that there is a 95% probability that the rejection rates of the N Z-tests will fall into the 95% confidence interval of $a \pm 1.96 \times \sigma_{Bernoulli, (a, 1-a)}$. As table 3.2 shows, among the 24 rejection rates, two rejection rates are significantly different from the 5% level of significance. Overall, these results suggest that the long-term performance evaluation framework is well-specified under the null hypothesis. The rejection rates appear to be independent of the length of time horizon. This implies that the long-term performance evaluation framework is free from the aggregation problems in many long-term performance tests in which the rejection of the null hypothesis is more severe when the time horizon is longer.

3.4.2 Power

The power of our long-term performance evaluation framework is primarily evaluated by artificially introducing a constant level of abnormal return to each of the available returns of event stocks. Following Barber and Lyon (1997) and Lyon, Barber, and Tsai (1999), we add ±0.42% (5%/12), ±0.84% (10%/12), ±1.26% (15%/12), and ±1.68% (20%/12) to each event stock's monthly returns to obtain induced levels of annual abnormal returns of ±5%, ±10%, ±15%, and ±20%, respectively. We then document the empirical rejection rates of the null hypothesis of the equalities between the bootstrap rejection rates of H_0: $PE_t \leq 0$ or H_0: $PE_t \geq 0$ for the 500 random pairs of event and pseudo samples.

In table 3.3, the power of our long-term performance evaluation framework is evaluated for the one-year horizon at the 5% level of significance for the bootstrap tests and the Z-tests. We chose the one-year horizon and the 5% level of significance because we would like to compare our framework with the existing methods whose

Table 3.3

Power of the Tests of the Null Hypothesis of the Equalities between the Rejection Rates of H_0: $PE_i \leq 0$ and H_0: $PE_i \geq 0$ for the 500 Random Pairs of Event and Pseudo Samples at One-Year Horizon

This table presents the power of the tests of the null hypothesis of the equalities between the rejection rates of H_0: $PE_i \leq 0$ and H_0: $PE_i \geq 0$ for the 500 random pairs of event and pseudo samples at one-year horizon during the sample period of July 1973 through December 1997. We first construct 500 random pairs of event and pseudo samples from the population of all NYSE, AMEX, and NASDAQ ordinary common stocks in the 1997 CRSP monthly files. Following Shumway (1997) and Shumway and Warther (1998), we adjust return observations for the delisting bias to mitigate survivorship bias. We then add ±0.42% (5%/12), ±0.84% (10%/12), ±1.26% (15%/12), and ±1.68% (20%/12) to each event stock's monthly returns to obtain induced levels of annual abnormal returns of ±5%, ±10%, ±15%, and ±20%, respectively. We then document the empirical rejection rates at the 5% level of significance of the null hypothesis of the equalities between the bootstrap rejection rates of H_0: $PE_i \leq 0$ or H_0: $PE_i \geq 0$ for the 500 random pairs of event and pseudo samples at different levels of induced abnormal returns. The power results in panel B and C are from Barber and Lyon (1997, table 6 and 8).

Induced Level of Abnormal Return (%)	-20	-15	-10	-5	0	+5	+10	+15	+20
Panel A. Power of the Long-Term Performance Evaluation Framework Based on the GLPM									
Z-Tests	88	65	38	17	6	19	47	81	97
Panel B. Power of t-Statistics Using CARs									
Reference Portfolios									
Size Deciles	99	98	82	35	5	32	87	100	100
Book-to-Market Deciles	99	98	84	36	4	30	86	100	100
Size/Book-to-Market Portfolios	99	98	84	36	5	32	87	100	100
Equally Weighted Market Index	99	97	76	25	5	39	91	100	100
Control Firm Approach									
Size-Matched Control Firm	98	89	59	19	5	17	56	86	98
Book-to-Market Matched Control Firm	98	88	58	21	6	17	54	86	97
Size/Book-to-Market Matched Control Firm	98	89	60	20	4	15	56	87	98
Fama-French Three-Factor Model α's	96	88	66	28	7	9	32	66	87
Panel C. Power of t-Statistics Using BHARs									
Reference Portfolios									
Size Deciles	97	91	77	42	9	11	57	96	100
Book-to-Market Deciles	97	92	79	44	10	9	55	95	100
Size/Book-to-Market Portfolios	97	92	78	44	10	10	56	96	100
Equally Weighted Market Index	96	90	72	35	8	14	66	97	100
Control Firm Approach									
Size-Matched Control Firm	91	76	45	14	5	14	44	74	91
Book-to-Market Matched Control Firm	91	75	46	17	4	13	42	72	90
Size/Book-to-Market Matched Control Firm	92	76	47	15	3	13	43	74	91

power issues are extensively examined by Barber and Lyon (1997) and Lyon, Barber, and Tsai (1999) who use one-year monthly returns in random samples at the 5% level of significance. As table 3.3 shows, the power of our long-term performance evaluation framework is quite similar to the power of the Fama-French three-factor model under the alternative hypothesis. However, the Fama-French three-factor model is shown by Barber and Lyon (1997) to yield negatively biased test statistics under the null hypothesis at one- and three- year horizons, while our framework is demonstrated in the previous subsection to be well-specified under the null hypothesis regardless of the length of time horizon.

We are interested in comparing our long-term performance evaluation framework to the control firm approach advocated by Barber and Lyon (1997). Barber and Lyon show that the control firm approach is well-specified under the null hypothesis. We find that our framework and the control firm approach have similar power when the control firm approach uses the buy-and-hold abnormal return (BHAR) metric which Barber and Lyon favor on conceptual grounds. Therefore, our framework is doing at least as well as the control firm approach in a statistical sense. The power of the control firm approach, bootstrapped skewness-adjusted *t*-statistic, empirical *p*-value, and our long-term performance evaluation framework is depicted graphically in figure 3.1. The similarity in power is evident in the figure.

More importantly, on theoretical grounds, as Loughran and Ritter (2000) argue, a positive benchmark controlling for size and book-to-market, such as the control firm approach of Barber and Lyon (1997), can only be used to test whether, other than size and book-to-market, another pattern exists. In this regard, we favor

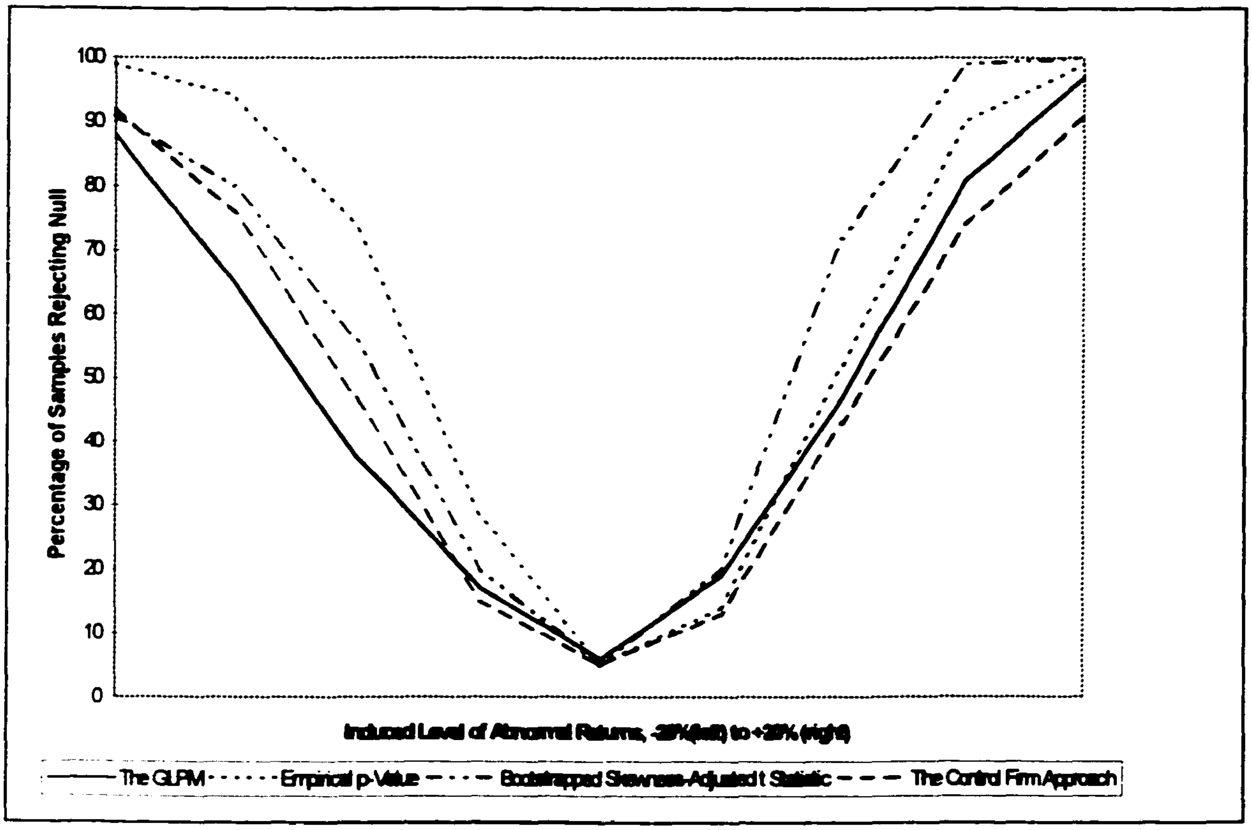

Figure 3.1 The power of the long-term performance evaluation framework based on the GLPM. The percentage of 500 random samples of 200 firms rejecting the null hypothesis at one-year horizon under the alternative hypothesis at various induced levels of annual abnormal returns (–20% to +20%) based on the GLPM. The power of the control firm approach, bootstrapped skewness-adjusted *t*-statistic, and empirical *p* values is from Lyon, Barber, and Tsai (1999).

the long-term performance evaluation framework based on the GLPM because the GLPM is a normative, equilibrium model. Similarly, although the use of the bootstrapped skewness-adjusted *t*-statistic and empirical *p*-value in Lyon, Barber, and Tsai (1999) improves the power, these two methods are subject to the same theoretical pitfall since they still rely on using size and book-to-market to generate expected returns. Therefore, although the power of the long-term performance evaluation based on the GLPM is slightly lower than the power of the bootstrapped skewness-adjusted *t*-statistic and the empirical *p*-value method, we believe that this framework is sound and competitive in a theoretical and conceptual sense.

In addition to the introduction of a constant level of abnormal return to each of the available returns of event stocks, the power of our long-term evaluation framework can be empirically evaluated by applying it to a known long-term anomaly. The idea is that if the framework is universally without power, then the framework will be unable to detect an anomaly even if the anomaly is robust. By the same token, if the framework is able to verify an anomaly for which the existing empirical results are mixed, we can be more confident about the power of the framework. In the following section, we chose IPOs to empirically evaluate the power of our long-term performance evaluation framework because the existing evidence regarding the underperformance of IPOs is mixed.

3.5 An Application: IPOs

After analyzing the statistical properties of our long-term performance evaluation framework, we apply this framework to the underperformance of IPOs

documented by Ritter (1991). We first review the underperformance of IPOs. The sample is then briefly discussed. Finally, the test results are reported.

3.5.1 The Long-Term Underperformance of IPOs

Ritter (1991) documents that IPO firms significantly underperform relative to non-IPO firms for three years after the offering date. Ritter argues that this underperformance is consistent with an IPO market in which investors are periodically, systematically overoptimistic about the prospects of IPO firms, and firms take advantage of these windows of opportunity. This IPO long-term underperformance has also been confirmed by Loughran and Ritter (1995) and Spiess and Affleck-Graves (1995).

Some studies argue that the long-term underperformance of IPOs is sensitive to the return metric. Brav and Gompers (1997) show that the underperformance disappears when the benchmarks control for size as well as book-to-market ratio. Brav and Gompers also find that when IPOs are value-weighted, instead of equal-weighted, five-year abnormal buy-and-hold returns shrink significantly. Fama (1998) argues that the underperformance is not robust to alternative methods and, therefore, the paradigm of market efficiency should not be abandoned.

Recently, Loughran and Ritter (2000) have demonstrated that the value-weighting scheme has low power in detecting abnormal returns when the event being studied is a managerial choice variable, such as IPOs. Furthermore, they show that IPO firms reliably underperform on both an equal-weighted and a value-weighted basis when the Fama-French (1993) three-factor regressions are run using factors that have been purged of IPOs to mitigate the benchmark contamination problem. In

short, the evidence regarding the underperformance of IPOs is mixed, and the debate on market efficiency appears to have no end in the near future.

3.5.2 The Sample

The IPO sample is from Ritter (1991). The sample comprises 1,525 IPOs that went public in the 1975-1984 period.[17] The post-event window includes the following 36 months where months are defined as successive 21-trading-day periods relative to the IPO date. For IPOs that are delisted in the post-event window, the data set uses the available return series and is adjusted for the delisting bias to mitigate survivorship bias (Shumway (1997), Shumway and Warther (1998)).

We construct a pseudo sample of 1,525 non-IPO stocks. For each IPO sample stock, a pseudo stock with the delisting return adjusted is randomly selected such that the pseudo stock has return observations that cover the beginning of the IPO sample stock's post-event window. This procedure is repeated to obtain a pseudo sample of 1,525 non-IPO stocks.

3.5.3 Test Results

We apply the bootstrap method to each of the 1,525 IPO sample stocks and to each of the 1,525 pseudo stocks. Table 3.4 presents the test results. In panel A, the rejection rates of H_0: $PE_t \leq 0$ and H_0: $PE_t \geq 0$ for the IPO sample are 1.38% and 5.64%, respectively, at the 1% level of significance. It is clear that IPOs tend to have long-term negative performance. Similarly, at the 5% level of significance, the rejection rate of H_0: $PE_t \geq 0$, 12.66%, is much higher than that of H_0: $PE_t \leq 0$,

[17] One of Ritter's (1991) sample, whose CRSP permanent number is 11077, is excluded because it has no return observations in the 1997 CRSP daily files.

Table 3.4
Test Results of the Null Hypothesis of the Equality between the Bootstrap Rejection Rates for the IPO and Pseudo Samples

This table presents test results of the null hypothesis of the equality of the bootstrap rejection rates for the IPO and pseudo samples. The IPO sample is from Ritter (1991). Following Shumway (1997) and Shumway and Warther (1998), the data set is adjusted for the delisting bias. For each IPO sample stock, a pseudo stock with the delisting return adjusted is randomly selected such that the return series of the pseudo stock covers the beginning of the IPO sample stock's three-year post-event window. The restricted sample exclude those stocks whose return series is shorter than 12 months. Another set of test results without the delisting returns adjusted is also reported. Monte Carlo realizations of bootstrap samples are then independently generated to obtain 1,000 bootstrap pricing errors for each stock. The empirical distribution of these bootstrap pricing errors is used to construct the bootstrap critical values. We then apply the Z-test for the equality between rejection rates to the IPO and pseudo samples.

		Theoretical Significance Level			
		1%		5%	
	# of	The Null Hypothesis			
	Stocks	H_0: $PE_t \leq 0$	H_0: $PE_t \geq 0$	H_0: $PE_t \leq 0$	H_0: $PE_t \geq 0$
Panel A. Ritter's (1991) Sample					
IPO Sample	1525	1.38	5.64	5.05	12.66
Pseudo Sample	1525	2.16	2.62	8.66	7.87
Z-Test Statistic			4.19*		4.36*
Panel B. Restricted Sample (A Minimum of 12-Month Return Observations)					
IPO Sample	1380	1.74	5.72	5.43	13.25
Pseudo Sample	1380	2.57	2.57	8.95	8.66
Z-Test Statistic			4.15*		3.86*
Panel C. Ritter's (1991) Sample without the Delisting Bias Adjustment					
IPO Sample	1525	1.38	4.26	5.38	11.48
Pseudo Sample	1525	2.17	2.62	8.68	7.86
Z-Test Statistic			2.48*		3.38*
Panel D. Restricted Sample without the Delisting Bias Adjustment					
IPO Sample	1380	1.88	4.49	5.87	12.25
Pseudo Sample	1380	2.57	2.57	8.88	8.62
Z-Test Statistic			2.73*		3.12*

* Significant at the 1% level of significance, one-tailed

5.05%. In contrast, the rejection rates of H_0: $PE_t \leq 0$ and H_0: $PE_t \geq 0$ for the pseudo sample are closer to each other, and not far from the specified levels of significance. Applying the Z-test for the equality between the rejection rates of H_0: $PE_t \geq 0$ for the IPO and pseudo samples, the test statistics, 4.19 for the bootstrap rejection rate at the 1% level of significance and 4.36 for the bootstrap rejection rate at the 5% level of significance, are significant at the 1% level for the Z-tests. These results suggest that the risk-return tradeoff in the IPO sample is different from the risk-return tradeoff in the random pseudo sample.

These results also not sensitive to whether we impose a minimum requirement for the number of monthly return observations on the IPO sample stocks. Among the 1,525 IPO sample stocks, 9.50% of them are delisted within 12 months subsequent to their IPO dates. Since the return observations are limited for these short-lived IPO stocks, one might question whether inferences based on these return observations are robust. To address this concern, we exclude those IPO sample stocks whose monthly returns series is less than 12 months and obtain a subset of 1.380 IPO sample stocks.[18] The test results on this restricted sample are reported in panel B of table 3.4. It is clear that the rejection rates for both the restricted IPO and the correspondent pseudo samples and the Z-statistics do not change much after imposing the 12-month-return restriction.

The conclusions favoring the window of opportunity argument of Ritter (1991) are also invariant to whether we take the delisting bias into consideration. In the existing studies, including Ritter (1991), Loughran and Ritter (1995), Spiess and

[18] Results based on different cutoffs are similar.

Affleck-Graves (1995), and Loughran and Ritter (2000), the delisting bias has not been adjusted. Consequently, one might argue that the test results presented in this study are driven by the missing return adjustment. To investigate this possibility, we exclude the adjusted delisting returns from our initial and restricted IPO and pseudo samples. The test results, shown in panel C and D, indicate that although the adjustment of the delisting returns makes the underperformance more severe, the initial and restricted IPO samples still appear to underperform even without the adjustment. The *Z*-test statistics, ranging from 2.48 to 3.38, are still significant at the 1% level of significance.

In short, the IPO long-term performance evaluation based on the GLPM is consistent with the previously documented underperformance (Ritter (1991), Loughran and Ritter (1995), Spiess and Affleck-Graves (1995), and Loughran and Ritter (2000)). The result is based on a statistic that is well-specified under the null hypothesis and is of power under the alternative hypothesis. The major difference between this study and the existing studies is that our long-term evaluation framework has a theoretical foundation for testing market efficiency because the statistic is based on a normative (equilibrium) asset pricing model.

3.6 Conclusions

In this essay, we analyze the long-term performance evaluation framework based on the GLPM. The GLPM is a normative (equilibrium) model with economic intuition and robustness. This study is in the line with the advocacy of Loughran and Ritter (2000) that long-term performance evaluation requires a normative (equilibrium) asset pricing model. In addition, this evaluation framework uses

bootstrap tests on an individual stock basis to mitigate the skewness problem of abnormal return distributions and avoid the aggregation problems in many long-term performance tests. Our simulation results show that this evaluation framework is well-specified and is of power.

We apply our long-term performance evaluation framework to Ritter's (1991) IPO sample. Consistent with Ritter (1991), Loughran and Ritter (1995), Spiess and Affleck-Graves (1995), and Loughran and Ritter (2000), we document that, according to the normative risk-return tradeoff of GLPM, IPO firms underperform in their three-year post-event windows. Since this is the first application of the normative (equilibrium) GLPM in long-term performance evaluation, it would be interesting to see how other long-term anomalies behave under this long-term performance evaluation framework. We hope that this study can motivate future research in this direction.

CHAPTER 4

CONCLUSIONS

This dissertation provides an empirical examination of the GLPM, in which loss aversion is intuitively incorporated into investors' portfolio decisions. In the GLPM equilibrium, the risk-return relation is based on the tradeoff between expected market-related gain and loss. In addition to its rich economic intuition, the GLPM is shown to be more robust than the mean-variance-based CAPM.

The bootstrap method is used to test asset pricing on an individual asset basis. First, the empirical power of the GLPM is examined for each sample stock. The testing framework used in this study imposes a higher testing standard and places no restrictions on the distribution of the statistic with the use of the bootstrap. The results show that the mean rejection rates of the GLPM are 2.03% and 7.04% for the 1% and 5% levels of significance, respectively, during the 1948-1997 sample period. The highest rejection rates in the 10 five-year test periods are 3.07% and 8.18% for the 1% and 5% levels of significance, respectively. Given the allowable level of the 1% and 5% levels of significance and the randomness in the level of the type-I errors, no more than 3% of sample stocks are mispriced. These results indicate that the GLPM appears to do a good job of explaining stock returns.

Second, based on these testing results, a long-term performance evaluation framework based on the normative (equilibrium) GLPM is proposed. This investigation is in the line with Loughran and Ritter's (2000) argument that long-term performance evaluation requires a normative (equilibrium) asset pricing model. This evaluation framework based on the bootstrap method is also capable of

mitigating the skewness problem of abnormal return distributions and avoiding the aggregation problems in many long-term performance tests. The specification of the long-term performance evaluation framework is evaluated using samples of randomly selected NYSE/AMEX/NASDAQ stocks and simulated random event dates. Simulation results show that the long-term performance evaluation framework based on the normative (equilibrium) GLPM is well-specified.

Overall, this dissertation provides empirical evidence in favor of the empirical power of the GLPM in explaining asset returns and in evaluating long-term performance. Because of its explanatory ability, it appears that the use of the GLPM in generating expected returns for empirical applications should be promising. Consequently, a future analysis of the use of the GLPM in determining the cost of capital for risky budgeting projects would be worthwhile.

In addition, based on the CAPM, researchers have identified a number of anomalies in asset prices. Empirical anomalies can arise from several sources, such as market frictions, false inferences of the risk-return relationship, inappropriate risk measures, and mispricing by market participants. The empirical results based on the GLPM, on the other hand, suggest that there are not many mispriced stocks. For future research, it would be interesting to apply the GLPM to re-examine these anomalies.

REFERENCE

Artzner, P., F. Delbaen, J. M. Eber, and D. Heath, 1997, Thinking coherently, *Risk* 10, 68-71.

Banz, Rolf W., 1981, The relationship between return and market value of common stocks, *Journal of Financial Economics* 9, 3-18.

Barber, Brad M., and John D. Lyon, 1997, Detecting long-term abnormal stock returns: The empirical power and specification of test statistics, *Journal of Financial Economics* 43, 341-372.

Basu, Sanjoy, 1977, The investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis, *Journal of Finance* 32, 663-682.

Bawa, Vijay S., and Eric B. Lindenberg, 1977, Capital market equilibrium in a mean-lower partial moment framework, *Journal of Financial Economics* 5, 189-200.

Benartzi, S., and Richard Thaler, 1995, Myopic loss aversion and the equity premium puzzle, *Quarterly Journal of Economics*, 73-92.

Berk, Jonathan B., 1995, A critique of size-related anomalies, *Review of Financial Studies* 8, 275-286.

Berk, Jonathan B., 2000, Sorting out sorts, *Journal of Finance* 55, 407-427.

Black, Fisher, Michael C. Jensen, and Myron Scholes, 1972, The capital asset pricing model: Some empirical tests, in: Michael C. Jensen, ed., *Studies in the Theory of Capital Markets* (Praeger, New York, NY).

Brav, Alon, and Paul A. Gompers, 1997, Myth or reality? The long-run underperformance of initial public offerings: Evidence from venture and nonventure capital-backed companies, *Journal of Finance* 52, 1791-1821.

Brown, Stephen J., and Jerold B. Warner, 1980, Measuring security price performance, *Journal of Financial Economics* 8, 205-258.

Brown, Stephen J., and Jerold B. Warner, 1985, Using daily stock returns: The case of event studies, *Journal of Financial Economics* 14, 3-51.

Daniel, Kent, and Sheridan Titman, 1997, Evidence on the characteristics of cross-sectional variation in stock returns, *Journal of Finance* 52, 1-33.

De Bondt, Werner F. M., and Richard Thaler, 1985, Does the stock market overreact? *Journal of Finance* 40, 793-805.

Dichev, Ilia D., 1998, Is the risk of bankruptcy a systematic risk? *Journal of Finance* 53, 1131-1147.

Dybvig, Phillip H., and Jonathan E. Ingersoll, Jr., 1982, Mean-variance theory in complete markets, *Journal of Business* 55, 233-252.

Efron, Bradley, 1979, Bootstrap methods: Another look at the jackknife, *The Annals of Statistics* 7, 1-26.

Efron, Bradley, and Robert J. Tibshirani, 1993, *An Introduction to the Bootstrap* (Chapman & Hall, New York, NY).

Fama, Eugene F., 1998, Determining the number of priced state variables in the ICAPM, *Journal of Financial and Quantitative Analysis* 33, 217-231.

Fama, Eugene F., 1998, Market efficiency, long-term returns, and behavioral finance, *Journal of Financial Economics* 49, 283-306.

Fama, Eugene F., and Kenneth R. French, 1992, The cross-section of expected stock returns, *Journal of Finance* 47, 427-465.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, , *Journal of Financial Economics* 33, 3-56.

Fama, Eugene F., and Kenneth R. French, 1996, Multifactor explanations of asset pricing anomalies, *Journal of Finance* 51, 55-87.

Fama, Eugene F., and James MacBeth, 1973, Risk, return and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607-636.

Ferson, Wayne E., and Campbell R. Harvey, 1999, Conditioning variables and the cross section of stock returns, *Journal of Finance* 54, 1325-1360.

Fishburn, P., 1977, Mean-risk analysis with risk associated with below-target returns, *American Economic Review* 67, 116-126.

Fisher, R. A., 1930, The moments of the distributions of normal samples of measures of departure from normality, *Proceedings of the Royal Society of London* 130, 16-28.

Greenblatt, J., 1997, *You can be a Stock Market Genius* (Simon and Schuster: New York, NY).

Hall, Peter, Joel L. Horowitz, and Bing-Yi Jing, 1995, On blocking rules for the bootstrap with dependent data, *Biometrika* 82, 561-574.

Hogan, William W., and James M. Warren, 1974, Toward the development of an equilibrium capital-market model based on semivariance, *Journal of Financial and Quantitative Analysis* 9, 1-11.

Jarrow, Robert A., and Dilip B. Madan, 1997, Is mean-variance analysis vacuous: Or was beta still born?, *European Finance Review* 1, 15-30.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65-91.

Judge, George G., William E. Griffiths, R. Carter Hill, and Tsoung-Chao Lee, 1980, *The Theory and Practice of Econometrics* (John Wiley & Sons, New York, NY).

Kahneman, Daniel, and Amos Tversky, 1979, Prospect theory: An analysis of decision under risk, *Econometrica* 47, 263-291.

Kahneman, Daniel, J. Knetsch, and Richard H. Thaler, 1990, Experimental tests of the endowment effect and the Coase theorem, *Journal of Political Economy*, 1325-1348.

Kanji, Gopal K., 1993, *100 Statistical Tests* (SAGE, London).

Keim, Donald B., 1983, Size-related anomalies and stock return seasonality, *Journal of Financial Economics* 12, 13-32.

Knez, Peter J., and Mark J. Ready, 1997, On the robustness of size and book-to-market in cross-sectional regressions, *Journal of Finance* 52, 1355-1382.

Kothari, S.P., and Jerold B. Warner, 1997, Measuring long-term security price performance, *Journal of Financial Economics* 43, 301-339.

Leland, Hayne E., 1999, Beyond mean-variance: Performance measurement in a nonsymmetrical world, *Financial Analysts Journal* 55, 27-36.

Lin, Ji-Chai, 1999a, A paradox of mean-variance preferences, working paper, Louisiana State University.

Lin, Ji-Chai, 1999b, Loss aversion, asset allocation, and gain-loss pricing, working paper, Louisiana State University.

Lintner, John, 1965, Security prices, risk and maximal gains from diversification, *Journal of Finance* 30, 657-675.

Litzenberger, Robert H., and Krishna Ramaswamy, 1979, The effect of personal taxes and dividends on capital asset prices: Theory and empirical evidence, *Journal of Financial Economics* 7, 163-196.

Lo, Andrew W., and A. Craig MacKinlay, 1990, Data-snooping biases in tests of financial asset pricing models, *Review of Financial Studies* 3, 431-468.

Loughran, Tim, 1997, Book-to-market across firm size, exchange, and seasonality: Is there an effect? *Journal of Financial and Quantitative Analysis* 32, 249-268.

Loughran, Tim, and Jay R. Ritter, 1995, The new issues puzzle, *Journal of Finance* 50, 23-51.

Loughran, Tim, and Jay R. Ritter, 2000, Uniformly least powerful tests of market efficiency, *Journal of Financial Economics* 55, 361-389.

Lyon, John D., Brad M. Barber, and Chih-Ling Tsai, Improved methods for tests of long-run abnormal stock returns, *Journal of Finance* 54, 165-201.

MacKinlay, A. Craig, 1995, Multifactor models do not explain deviations from the CAPM, *Journal of Financial Economics* 38, 3-28.

Mao, J., 1970, Survey of capital budgeting theory and practice, *Journal of Finance* 25, 349-360.

Markowitz, Harry M., 1959, *Portfolio Selection: Efficient Diversification of Investments* (Wiley & Sons, New York, NY).

Mossin, J., 1966, Equilibrium in a capital asset market, *Econometrica* 34, 768-783.

Reinganum, Marc R., 1982, A direct test of Roll's conjecture on the firm size effect, *Journal of Finance* 37, 27-35.

Ritter, Jay R., 1991, The long-run performance of initial public offerings, *Journal of Finance* 46, 3-27.

Sharpe, William F., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *Journal of Finance* 19, 425-442.

Sharpe, William F., 1998, Morningstar's risk-adjusted ratings, *Financial Analysts Journal* 54, 21-33.

Shumway, Tyler G., 1997, The delisting bias in CRSP data, *Journal of Finance* 52, 327-340.

Shumway, Tyler G., and Vincent A. Warther, 1998, The delisting bias in CRSP's Nasdaq data and its implication for the size effect, *Journal of Finance*, forthcoming.

Speiss, D. Katherine, and John Affleck-Graves, 1995, Underperformance in long-run stock returns following seasoned equity offerings, *Journal of Financial Economics* 28, 243-268.

Tversky, Amos, and Daniel Kahneman, 1986, Rational Choice and the framing of decisions, *Journal of Business* 59, s251-s278.

Tversky, Amos, and Daniel Kahneman, 1991, Loss aversion and riskless choice: A reference dependent model, *Quarterly Journal of Economics*, 1039-1061.

Tversky, Amos, and Daniel Kahneman, 1992, Advances in Prospect Theory: Cumulative representation of uncertainty, *Journal of Risk and Uncertainty* 5, 297-323.

VITA

Kevin C.H. Chiang received a diploma in mining engineering from National Taipei Institute of Technology in 1986. In 1992, after working as a security analyst for four years, he matriculated in Mankato State University's Master of Business Administration Program and received a Master of Business Administration degree in 1994. In 1996, he entered the doctoral program in Business Administration at Louisiana State University and will complete his doctoral degree in the summer of 2000.

DOCTORAL EXAMINATION AND DISSERTATION REPORT

Candidate: Kevin Chun-Hsiung Chiang

Major Field: (Finance) Business Administration

Title of Dissertation: Essays on the Gain-Loss Pricing Model

Approved:

Major Professor and Chairman

Dean of the Graduate School

EXAMINING COMMITTEE:

Date of Examination:

June 23, 2000