PERFORMANCE PERSISTENCE IN MANAGED FUTURES FUNDS

By

JOHN P. TOWNSEND

Bachelor of Science
New Mexico State University
Las Cruces, New Mexico
1992

Master of Science
New Mexico State University
Las Cruces, New Mexico
1994

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
July, 1998

PERFORMANCE PERSISTENCE IN MANAGED

FUTURES FUNDS

Thesis Approved:

Wade Brorsen

Thesis Adviser

Raymond Joe Schatzer

Kim B. Anderson

Jim Krehbiel

Wayne B. Powell

Dean of the Graduate College

ACKNOWLEDGMENTS

I wish to express my appreciation to my major advisor, Dr. Wade Brorsen, for his infinite patience, guidance, and support. I also want to thank Dr. Kim Anderson for his support and guidance and most of all, his ability to make me laugh when I really needed to. I also want to thank Dr. Tim Krehbiel for his support and insightful suggestions to improve this study and make it more relevant. A special thanks go to Dr. Joe Schatzer for allowing me to complete my teaching practicum under his guidance and for giving me the opportunity to do what I love to do the most, teach.

I cannot forget to thank those who picked me up when I wanted to leave, helped me with more than one econometrics problem and made me realize that I am blessed to count them as some of the finest people I have ever met and I am proud to call them my friends. First, to John and Andrea Lehr, two wonderful people who took this bachelor in and let him become a part of their lives, thank you. To Chris, who put up with more grief and laughed harder at my tee-offs than anyone else, but who also reminded me not to take myself too seriously, thanks, buddy. To Wes, as good an officemate as anyone could have, and finally, to Kurt, who always stood behind me and wouldn't let me get down. The man who huh,huh,huh, never, ever, let me down. Thanks.

Finally, I want to thank the Department of Agricultural Economics for supporting me these three years and giving me the opportunity to obtain this degree.

TABLE OF CONTENTS

LIST OF TABLES

NOMENCLATURE

CAPM	capital asset pricing model
CTA	commodity trading advisor
EGR	Elton, Gruber, and Rentzler (1987)
EGLS	estimated generalized least squares
FMDR	Foundation for Managed Derivatives Research
GLM	generalized linear model
IKZ	Irwin, Krukmeyer, and Zulauf (1992)
IML	Interactive Matrix Language
SML	security market line
α_{per}	calculated intercept from 1 year performance period
α_{sel}	calculated intercept from 4 year selection period
$(\alpha/\sigma)_{per}$	calculated Sharpe ratio from 1 year performance period
$(\alpha/\sigma)_{sel}$	calculated Sharpe ratio from 4 year selection period
$mean_{per}$	mean return calculated from 1 year performance period
$mean_{sel}$	mean return calculated from 4 year selection period
R_F	risk-free rate
K_M	market rate of return
r_s	Spearman's rank correlation

w_i	weight of residual

CHAPTER 1

INTRODUCTION

Problem Statement

Managed futures funds, both in the number of funds and the assets under management, have been growing rapidly. Irwin and Brorsen (1985) ascribe the genesis of the large managed futures industry to the mid-1970's although the first futures fund was considerably earlier. The growth has continued up to the present with over $25 billion invested in managed futures (McCafferty 1994). Managed futures have some similarity to stock mutual funds. Both managed futures funds and stock mutual funds build a portfolio of their respective securities, for managed futures funds this consists of a portfolio of futures contracts and other assorted securities. In stock mutual funds the portfolio consists of assorted stocks and varying amounts of other assorted securities.

There are many reasons why investors choose to invest in mutual funds. Many of these reasons for choosing to invest in mutual funds are equally applicable to managed futures funds: customer service, low transaction costs (in the form of low overhead and low commissions but not to be interpreted as low fees charged by the fund manager), diversification, and professional management (Gruber 1996). Low transaction costs differ from fees charged by fund managers in that transaction costs refer to commissions and overhead associated with mutual funds. The fees charged by fund managers are fees charged for the perceived management ability of the fund manager and will be incurred even if no trading occurs.

A new article by Goetzmann and Brown (1997) argues that hedge funds exist to exploit market inefficiencies. Hedge funds have fewer restrictions on trading than mutual funds. Hedge funds can short sell, arbitrage, and trade derivatives. Funds exist because it is the cheapest way for the fund manager to raise capital. Since different funds exploit different inefficiencies, different levels of performance can be expected. The returns could be a function of dollars under management. If the amount of dollars under management is small relative to other funds, it may be easier for the fund manager to exploit inefficiencies. As the dollars under management flows into a superior performing fund, the fund manager may have more difficulty in exploiting inefficiencies due to the larger amount of dollars under management. Fund managers may also increase management fees which offset the returns (part of the returns are consumed in the manager's fees).

Teweles and Jones (1987) identify several reasons why investors turn their money over to professional fund managers but state, "the primary reason is probably the perceived expertise of the professional manager." (Teweles and Jones, p. 255) There are several other reasons cited, such as the ability of funds to diversify the portfolio into many futures markets and through this diversification, reduce the risk of the fund.

With growth in the managed futures market, the question now becomes which fund or funds should be invested in? In the fund selection process undertaken by investors, a valid question to be answered is, do funds that have done well in the past tend to also do well in the future? The existence of performance persistence with managed futures funds has been a source of debate. Funds are said to exhibit performance persistence if the funds that do best in one period tend to do as well or better in the next period. If persistence exists, then selecting which funds to invest in would include the historical returns of each fund as well as other factors.

The issue of whether stock mutual funds exhibit performance persistence is also controversial. Whether or not performance persistence exists depends on which study is being evaluated. Volkman and Wohar (1995) found performance persistence in mutual funds and attempted to break down this fund persistence into measurable parts. Elton, Gruber, and Blake (1996) found that in risk-adjusted mutual fund returns, the past return does contain information about future performance of that fund. Carhart (1997) was able to explain away almost all predictability of returns as factors of strategies and fees. Gruber (1996) found that performance persistence not only existed, but also commented on the strength of the persistence.

It is important to note that in mutual funds, it is possible to purchase stocks with the intention of using a buy-and-hold strategy. This strategy, with a well diversified portfolio, may yield positive annual returns as long as the trend for those stocks is upward. The stocks could be held indefinitely and yield positive returns. This differs from a managed futures fund in that futures funds may practice short term hold-and-buy strategies only until the delivery date of the futures contract, at which time the contract must be offset or delivery made or taken. Kolb (1992) found that for most commodities, the expected return of both a buy and hold or sell and hold strategy would be zero because prices do not tend to be higher or lower the longer the time until expiration. This lack of a trend in prices being higher or lower makes taking either position, buy and hold or sell and hold, have zero returns.

An important point to raise is that futures funds must use professional management because a buy-and-hold strategy in futures would generate substantial transactions costs (due to rollovers) and has been shown to be a losing strategy (Lukac, Brorsen, and Irwin, 1988). The necessity of active management may allow easier identification of performance persistence, as any positive return requires professional management, as opposed to positive stock returns which can be obtained by using a buy and hold strategy.

However, Schwager (1996) found in his review of literature of performance persistence in managed futures funds, inconclusive evidence that the best performing funds can be predicted. Zweig (1996) pointed out that using Gruber's method of performance persistence (picking the top 10% of funds each year from a fixed list of 227 funds) would only slightly outgain (+$2156) a buy and hold strategy using an index Gruber developed to measure the risk of the whole market.

The question posed is does performance persistence exist in managed futures funds? By using past performance as the only selection criterion, can the superior performing fund or funds be selected? This study will attempt to answer the questions by determining if historical returns are a predictor of future performance by using regression analysis and the non-parametric analysis described by EGR's method having corrected for heteroskedasticity (Gruber, et al 1996). The study was expanded to include measures of risk / return in the form of an adjusted Sharpe Ratio.

Managed futures funds are similar in some aspects to mutual funds in that a pooling of funds from different investors are held united and traded as one pool or fund. For this study, the terms fund and pool are interchangeable. These pools or funds may be managed and traded by one commodity trading advisor (CTA) or a team of CTA's, depending on the choice of the pool operator. The CTA determines, according to their own preferred trading style, what percentage of the funds under their management will be used to cover margins incurred from positions in the futures market. This is defined in this study as the "leverage" of the fund. For many futures funds, margins necessary for the holding of positions have been put in the form of U.S. securities instead of cash. The futures fund puts up U.S. securities whose current market value is equal to or greater than the margin requirement. The most common example being Treasury Bills. This allows the funds necessary for margin maintenance to earn interest. The remainder of the funds in the pool are also usually invested in U.S. Treasury Bills.

The funds are classified as either public or private. Public funds are offered to the public and that generally have smaller minimum investment requirements than private funds. These funds can be advertised, and the fund manager must provide a prospectus to the potential investors. Public funds must be registered with the Securities and Exchange Commission (SEC) and the Commodity Futures Trading Commission (CFTC).

A private fund also consists of pooled funds but is not available to the general public. This means the private fund cannot be advertised, and solicitations for the fund to the public cannot be made by fund managers. Private funds are not required to be registered with the Securities and Exchange Commission. CTA funds are distinguished by being managed by a single account manager. The CTA data include returns earned from trading for public and private futures funds.

Objective of the Study

The general objective of this study is to determine if performance persistence exists in managed futures funds. The specific objectives of the study are:

1) To determine if the weighted monthly mean returns for each fund across time are significantly different from each other for public funds, private funds, and CTAs.

2) To determine the size and power of prior methods that used a nonparametric method consisting of Spearman Rank correlation relative to a new approach using regression which adjusts for changes in aggregate performance of the funds and heteroskedasticity.

3) To determine if the historical mean returns, intercepts, or adjusted Sharpe ratios of a fund can predict the future performance of that fund.

Outline of Procedures

For each of the objectives, the data are separated into each fund type: CTA data, public fund data, and private fund data (Laporte Management ltd., unpublished data).

For the first objective, individual fund monthly returns are regressed against an overall monthly mean return calculated from all funds. Slopes and intercepts are allowed to differ by fund. The hypothesis tested is whether the intercepts calculated for each fund in the regression are all equal. This hypothesis tests if any fund has a significantly different intercept from the other funds. This significantly different intercept is analogous to the "skill component that cumulates over time" suggested by Goetzmann and Ibbotson (1994).

The second objective will be accomplished with a Monte Carlo study where simulated returns are generated with different values and assumptions (SAS, 1990). The Monte Carlo study will test if the nonparametric methods used by Elton, Gruber, and Rentzler (EGR) can in fact detect performance persistence from data sets generated by Monte Carlo methods. Using these Monte Carlo methods, one data set was generated with performance persistence present and the other data set was generated with no performance persistence present. By evaluating the non-parametric methods used by EGR against both generated data sets, the size and power of EGRs methods can be determined.

The final objective is accomplished by calculating measures of persistence for the selection period and the performance period. Spearman's coefficients are calculated between the rank of the persistence measure calculated in the selection period and the rank of the persistence measure calculated in the performance period. The Spearman coefficients use a test statistic calculated for each selection/performance period to determine if the ranking (based on return) of the funds between the selection and performance period is significant. The final evaluation method uses the method used by Goetzmann and Ibbotson (1994) in determining if persistence is present in mutual funds. A two-way contingency table of ranked fund returns of various selection and performance periods is presented to determine if funds that were winners in the selection period continue to perform well in the performance period due to skill of the manager.

Organization of the Study

Theoretical basis and the review of literature are contained in Chapter II. Chapter III presents the objectives in order by objective and method used for that objective. Chapter IV presents the results from each objective. Chapter V presents the conclusions.

CHAPTER II

DEFINING AND IDENTIFYING PERFORMANCE PERSISTENCE

Fund Performance

The efficient market hypothesis has three forms as put forward by Fama (1970). The weak form of market efficiency indicates that prices reflect all information available from past prices (Teweles and Jones 1987). This form is based on the premise that all information related to past prices and trading, such as volume and open interest, if indeed containing predictive value, is already built into the current price. This argument usually brings a chorus of criticism from those traders using technical trading systems, which the theory says would be ineffective if the market price reflected all information from past prices.

The semi-strong form deals with the speed with which price adjustment occurs given a release of information. This form argues that all published information is already reflected in the price (Teweles and Jones 1987). Information in the form of market reports, government reports, earnings reports, and the like already have the information built into the price. Using this argument, to successfully trade, a trader would have to have access to a source of information not readily available to the whole market or obtain that information prior to release to the market.

The strong form of the hypothesis maintains that prices reflect all information that can be acquired (Teweles and Jones 1987). This form is concerned with the availability of information to all markets participants, in that even information researched by imaginative researchers, is already included in the price.

Grossman and Stiglitz (1980) argued that it was impossible for markets to be informationally efficient. They proposed a model to refine the efficient markets notion that allows for traders to profit. The model is based on the premise that prices reflect information from informed traders, but only partially. They argue that informed traders are compensated for the risk that they take. Informed traders profit by taking "better" positions than their uninformed counterparts.

Jensen (1969) defines an efficient market as one where all past information available is reflected in the current price. Jensen argues that if security prices actually follow the strong form of the efficient market hypothesis, then traders and fund managers can not earn above average returns by trying to predict future prices based on past information. The only way to earn superior returns is to be the first with a new piece of information not available to all traders. This appears to follow Grossman and Stiglitz's theory of small disequilibriums in markets caused by differences in information that allow some to earn superior returns.

These forms of the efficient market hypothesis ask some interesting questions concerning whether performance persistence can exist or not. The first question, and most relevant question asked is, is performance persistence, by definition, an inefficiency in the markets? Since performance persistence reflects a fund manager's superior ability, or the ability to consistently outperform other fund managers, by definition this is an inefficiency in the market. However, this inefficiency would be short lived. A fund manager that identifies a superior approach will be recognized by investors and have more money flow into that manager. This increase in dollars under management would cause a decrease in the managers returns to the equilibrium level. This reduction in returns follows the theory that no above normal returns can be consistently earned by managers.

The information gathered by the trader may be difficult to obtain or process. The trader has separated themselves from the rest of the market by establishing a monopolistic hold on some form of information.

Another form that market inefficiency may take is the timing and availability of information. When a manager receives the information may allow for small inefficiencies to occur in the market. The managers may actively take advantage of those inefficiencies in information delivery. This inefficiency may quickly disappear as instantaneous information delivery is now becoming the norm around the world, making information delivery inefficiencies less frequent. However, most managed futures traders are not information traders, they are technical traders. Beja and Goldman (1980) suggested the theory that "the trendists demand is not based on a security's fundamental prospects". (Beja and Goldman, p. 245) Trendists do respond to the price adjustments in the market brought about by fundamental traders. Trendists trade based on the information used by fundamental traders.

Fama postulates that the long-term returns from trading in the futures market using one CTA or several, as in a private or public fund will have gross profits of zero since the empirical research seems to support the efficient market hypothesis. The expected return from such strategies is zero because Fama assumes no transaction costs. If transactions costs are included, then net profits will be negative. However, even if all traders do not have all the information, according to Fama, this allows for the efficient market hypothesis to hold since Fama stated "the market may still be efficient if a sufficient number have the information".

Assuming the market is truly efficient, investors, upon hearing of the superior performance of the fund manager, would flood the manager with new funds to be managed. This influx of capital would cause the returns of the fund manager to go down and the returns from the fund to go down as well.

Jensen points out that the "loading charge" is basically a sales fee that the fund manager charges for the service of managing funds. The data sets tested do not include information regarding loading fees. Is it not reasonable to assume that fund managers who were successful would raise their loading charge and fees to reflect their "better" management style? Would managers increase these fees to capture the additional returns until the return to investors was that of the risk free rate?

Following this line of theory, performance persistence would not exist. If investors could accurately predict a fund's future returns based on its past returns, then the investor could effectively predict which funds will produce the highest future returns. Having this ability, investors would simply select the fund or funds with the highest historical returns and obtain positive gross profits which conflicts directly with Fama's theory. However, there are several weaknesses in this line of reasoning. First, managed futures are a relatively new investment and so the information investors need to evaluate returns may not have been available. Second, the earning preferences of the CTAs may cause them to choose different fee structures. Finally, capital may be constrained for various reasons.

If performance persistence exists, then historical returns can help predict future returns. For this study, modeling the returns involves a model analogous to the Capital Asset Pricing Model (CAPM). Hirt and Block (1993) defined the CAPM as a model by which an asset is valued based on its risk characteristics. This model creates a new investment variable, it being the risk-free rate and combining it with the efficient frontier concept. The risk-free rate is defined as the lowest risk security with zero default risk. U.S. Government securities are considered risk-free. The efficient frontier is a linear representation of risk-return possibilities for different investment portfolios.

The model used in this study is similar to the security market line (SML) which is used to express the trade off between return and risk for an individual stock. For this study, the SML provides the tradeoff for an individual fund. The general form of the SML:

(1) $$k_i = R_F + \beta_i(K_M - R_F)$$

where k_i is the anticipated return based on this formula, R_F is the risk-free rate of return, β_i is the measure of the volatility of the security with respect to the market in general, and K_M represents the market rate of return.

The model proposed for the study attempts to model the stochastic process:

(2) $$r_{it} = \alpha_i + \beta_i \bar{r}_t + \epsilon_{it} \quad i = 1, \ldots, n \qquad t = 1, \ldots, T$$

$$\epsilon_{it} \sim N(0, \sigma_t^2)$$

where r_{it} is the return of fund (or CTA) *i* in month *t*, parameter β_i represents differences in leverage, r_t is average fund returns in month *t.* β_i allows each fund to have a different variance which is consistent with research. For this study, the term "leverage" represents the percentage of dollars in the fund devoted to margins. Leverage acts as a risk indicator since it reflects the percentage of funds tied to margins as opposed to funds invested in “safer” securities, specifically government securities, T-Bills, bonds, etc. Fund returns are generated as a function of the risk-free rate, the leverage the fund uses, and the monthly mean fund returns.

If persistence did not exist, then based on the model above, a regression of the returns should reveal the intercept of each fund to not be significantly different from the intercept of all the other funds. The intercepts should be a proxy for the risk-free rate. Although (2) is not exactly a CAPM, it does have components that form a basic structure that follows the thinking of the CAPM. This being that returns are a function of the risk-free rate and a measure of risk for each fund. Even though the funds may have different leverage amounts, the risk-free rate should be uniform across all funds. If the intercepts are different across years for different funds then that could be interpreted as the intercept is not just the risk-free rate. The intercept is instead is:

$$\alpha_i = f(R_F\ ,\ \rho_i) \qquad (3)$$

where ρ_i is some factor unique to that fund. It has been debated to exactly what ρ_i should represent. Two major theories have been put forward. One being that the factor is the management style of the CTA running the fund. Management style refers to several factors used by the CTA in managing the fund including leverage, commodities traded, or trading systems used. The other theory argues that it reflects the transaction costs of the fund. For this study, since persistence is be tested, it will be assumed that the management style is what will affect the intercept.

Previous Studies in Mutual Funds

Performance persistence has been researched more extensively in the field of mutual funds than in the field of managed futures funds. This is expected as mutual funds have existed longer than managed futures funds. Carhart (1997) tested for performance persistence in mutual funds using three different performance measurement models including the CAPM and a four-index model designed by Carhart. He determined that there was only slight evidence of skilled or talented fund managers and that persistence only lasts one year.

Zweig (1996) states that no one has proved persistence exists in mutual funds. However, citing Gruber, he points out that Gruber found by buying last year's hottest funds, he could beat out the market index average for returns. He also showed that trading on yearly winners in the mutual fund market does increase returns, but only slightly, and any gain is lost to paying transaction costs.

Kahn and Rudd (1995) used "style" analysis. From the results, it was determined that in both equity and fixed income funds, investors should not base selection of a trading advisor solely on historical performance. However, persistence was found to be present in fixed income funds, but past performance should not be the sole determinant of what fund to invest in.

Elton, Gruber, and Blake (1996) found that in mutual funds, historical performance does contain information about future performance and the intercepts, referred to from this point on as the α's, from 1 and 3 year selection periods convey information about future performance. They also found, like Carhart, that for 1 year performance periods, the prior year's performance appears to contain the most relevant information.

Volkman and Wohar (1995) chose to not only study if persistence exists, but if so, what are the components of that persistence. They identified three systematic components that influence persistent fund performance. These are past performance, fund goal, and management fee (Volkman and Wohar 1995).

Khorana and Nelling (1997) devoted a small portion of their article to the study of performance persistence in a specific mutual fund type called sector funds. Two methods were used, a cross-sectional comparison for performance across sectors and a comparison to a known index, in this case, the S&P 500. A nonparametric runs test was performed to check for sector fund manager persistence. A run was defined as an uninterrupted sequence of good-performance or bad-performance months (Khorana and Nelling 1997). The tests found that among 123 sector funds, only 15 exhibited persistence, most of which were under performing, that is, those that did poorly continued to do poorly (Khorana and Nelling 1997).

Goetzmann and Ibbotson (1994) addressed several issues in their paper on mutual fund persistence. They stated that "The investment performance of an individual mutual fund is likely to contain both a skill component and a noise component. The skill component would cumulate over time, while the noise component would be serially independent so that its average would tend toward zero over time." As compared to other studies, Goetzmann and Ibbotson chose to compare fund returns to each other, rather than to an absolute benchmark. Results showed significant performance persistence in multiple selection periods including two-year selection and two-year performance periods, persistence in one-year selection and one-year performance periods, and when funds were differentiated according to the "style" of the fund. Monthly returns were also tested to determine if a month's rank is related to last month's rank. Persistence was found in raw return data and a measure of risk called the "Jensen measure" pulled from Jensen's article, which is the alpha from the CAPM empirical analogue (Goetzmann and Ibbotson 1994). The article points out that to effectively select winning fund advisors, a long selection period should be chosen to eliminate noise in the performance data, but not so long as to allow the fund manager to change the trading system being used.

As with Goetzmann and Ibbotson's article, the potential for survivorship bias exists in our data set. Some of the commodity funds were closed down, most likely poor-performing funds. However, the survivorship bias in our study is partially mitigated by the same reason that it was in Goetzmann and Ibbotson's article. The bias is partially mitigated by the fact that our study compares survivors to other survivors relative performance (Goetzmann and Ibbotson 1994). Fund performance is not measured against an absolute benchmark, but rather against other funds.

Phelps and Detzel (1997) use the same methodology as Goetzmann and Ibbotson but expand it to include a risk control. They contended that persistence found in prior research was a result of insufficient risk controls (Phelps and Detzel 1997). They argued that any persistence found may be a form of macropersistence, or persistence in broad equity classes, rather than micropersistence, or the skill of an individual trader. Using indices to evaluate whether macropersistence or micropersistence was responsible for the prior results. Using one, two and three year selection periods, they found no significant persistence in the data and concluded "It does not appear that there is a reliable strategy for selecting funds expected to have superior future performance, other than to avoid funds with high expense ratios." (Phelps and Detzel, p. 67)

Gruber, in his 1996 presidential address, not only stated that persistence existed, but that he was amazed by the strength of persistence. Gruber stated he had found persistence and that sophisticated investors money flowed into and out of funds based on indicators of future performance.

Overall, the research seems inconsistent with some studies using the same methodology finding different results. Goetzmann and Ibbotson found significant persistence using two-year selection and performance periods. Phelps and Detzel used the same methodology but expanded it to include a risk control and found no significant persistence in the data. Carhart found little evidence of skilled managers. Khorana and Nelling no evidence of persistence in sector funds. Elton, Gruber, and Blake found strong evidence of persistence. Kahn and Rudd found that persistence was present. Overall, the evidence is inconclusive, but does seem to favor short-run persistence.

Previous Studies in Managed Futures Funds

There have been a few studies of performance persistence in managed futures funds including Elton, Gruber, and Rentzler (1987), Edwards and Ma (1988), Irwin, Krukemeyer, and Zulauf (1992), McCarthy, Schneeweis, and Spurgin (1997), and Schwager (1996). Elton, Gruber, and Rentzler (EGR)(1987) found that superior performing commodity funds generally could not be selected based on past performance of that fund. The exception being risk level of a fund, measured by EGR as the standard deviation of a fund. EGR found that standard deviation of a fund appeared to be a good indicator of the relative riskiness of a fund. Schwager also reviewed several studies that were concerned with determining if performance persistence is present in returns and risk levels. These include EGR, Edwards and Ma, Irwin, Krukemeyer, and Zulauf (IKZ), Irwin, McCarthy, and Irwin, Zulauf, and Ward.

Edwards and Ma limited their study to prepublic pool returns and postpublic pool returns with a goal of determining if the CFTC should change the regulatory fashion with which public fund returns are disclosed. They bring up the point if the past performance information is useless, should the CFTC continue to mandate its publication? The methodology involved regressing two years of post-public returns against three years of pre-public returns, the fee structure used by that particular pool, conditions during the futures market during the post-public returns, and an error term. With the measure of persistence being a significant positive relationship between the pre-public returns and the post-public returns (Edwards and Ma 1988). The prepublic mean monthly return across all funds was 4% while the postpublic mean monthly return was -.1%. Their conclusion addressed a different issue than this study, the issue of performance persistence between pre-public pool returns and post-public pool returns. Edwards and Ma studied if performance persistence was present between a fund's returns before it went public and the returns of that fund after it went public. The results showed no performance persistence was found between pre-public pool returns and post-public returns.

Prospectuses are required by law to include the statement "Past performance is not indicative of future performance." Edwards and Ma asked if this is an accurate statement. In fact, this a lawyerly statement that is mostly true for several reasons. First, it is important not to confuse relative fund performance with absolute fund performance. Relative performance of a fund compared to other funds does not mean that the absolute returns earned by that fund in the past will have predictive power as to the absolute returns in the future. Second, the problem of selectivity bias pointed out by Edwards and Ma also must be addressed. Selectivity bias is the when funds are selected on the basis of past returns when past returns explain only a small portion of the variation in a return and random error explains most of the variation in a return. If most of the variation in the fund return is not explained by the variable used to select which fund to invest in, then logically, the future returns will not be accurately predicted by looking at past fund returns.

McCarthy, Schneeweis, and Spurgin (1997) concentrated on CTA's combining to form public commodity funds. They found that pro forma unadjusted historical returns from each CTA tend to overestimate post-offering returns from that fund. This would seem to support the results found by Edwards and Ma. However, when risk is included in the returns, past returns are more accurate in predicting later returns than when risk is excluded from returns. This article also confirmed what several other studies have found such as EGR and others, in that past risk is a good indicator of future risk of the fund.

Irwin, Krukemeyer, and Zulauf compared the correlation between a pool's return in one year and its return in the subsequent year over the time period 1979-1989 (Irwin, Krukemeyer, and Zulauf). The correlations were calculated for pool performance for 4 pool groups: all pools, the top third performing funds, the middle third performing funds, and the bottom third performing funds. Standard deviations and Sharpe ratios were also compared from one year to the subsequent year. When all pools are considered, they concluded that only the standard deviation had a sufficiently large correlation coefficient (.451) to be economically meaningful. When pools are stratified into thirds, the correlations are slightly larger, but IKZ still conclude that "it is debatable whether any strategy to select public commodity pools can be used to obtain an economically meaningful increase in performance." (Peters, p. 422)

EGR made one adjustment in their analysis of performance persistence to account for the possibility that all fund variances were not the same when using returns as a predictor of future returns. One of the measures of performance used by EGR is a Sharpe Ratio, which is calculated using the standard deviation of the fund and therefore allows for different fund variances. They used a nonparametric method to test for performance persistence. This nonparametric method consisted of the Spearman Rank correlation. This nonparametric method consisted of comparing the relative rank of the annual return of a fund during a historical period (referred to from this point on as the selection period) to the relative rank of the same funds return during the following year (from this point on referred to as the performance period).

Some evidence to support the argument that performance persistence exists was found in the form of risk levels and mean returns from year to year. However, they chose to conclude that there was no evidence of performance persistence in mean returns. Schwager comments that EGR appears to have already made up their minds or simply ignored any evidence that contradicted their opinion. EGR misidentified a problem, not recognizing the heteroskedasticity in the data, and used a method that has virtually no power to reject the null hypothesis of no predictability in mean returns.

However, the prior literature may have misidentified a statistical problem, which is actually heteroskedasticity. To confirm that each of the raw data sets does violate the homoskedasticity assumption, a test for heteroskedasticity was performed on the public, private, and combined CTA data sets. Each of the funds use differing amounts of leverage, some use high margin, highly volatile commodities while others use less volatile, low margin commodities. Knowing this, the assumption of homoskedasticity does not hold. Even though this assumption does not hold, prior studies still used nonparametric methods. Nonparametric methods can not correct for heteroskedasticity caused by different amounts of leverage between funds except for the Sharpe Ratio used in prior studies. This paper takes a different approach from the prior studies by assuming heteroskedasticity among different funds and correcting for it by using regression analysis rather than nonparametric methods.

Prior literature used methods devised by Elton, Gruber, and Rentzler (EGR). These methods were found to have no power to reject the null hypothesis of no predictability (Grossman, 1987) when the assumptions of EGR's method were true. This would bias the results to show no significant persistence even when persistence was substantial. The power problems also include small sample size and a problem with short-run negative autocorrelation. A Monte Carlo study of the methods of EGR will determine if Grossman's criticisms of their methods are justified. Generating data sets with various conditions, nonparametric methods can be evaluated for ability to identify when persistence is present.

Correlations between funds are important because if one fund with less leverage did well when funds with larger amounts of leverage are doing poorly, a negative correlation exists. If both groups do well at the same time, a positive correlation exists. Neither EGR or Schwager account for different levels of leverage used by different fund managers. Failure to account for the different levels of leverage, the two methods argue that the fund managers use the same trading systems or that the systems differences are negligible, ignoring the heteroskedasticity problem.

CHAPTER III

PROCEDURES

This chapter presents procedures used in this study. The methods used to transform the data and to check for data accuracy are presented. The methods used to test the homoskedasticity and normality assumptions are explained. Next, the procedures used to conduct hypothesis tests on monthly mean returns are presented. The procedure for the Monte Carlo analysis of the nonparametric method employed by EGR is presented as is the procedure for using the same nonparametric method on the futures fund data. Finally, the nonparametric persistence tests used by Goetzmann and Ibbotson are described.

Data

The data are grouped by whether the returns are for a public fund, a private fund, or a commodity trading advisor (CTA). The data begin in 1978, but few funds were trading at that time. Public funds are defined as funds that were publicly offered to any investor with sufficient capital. The minimum capital requirement for these funds is generally less than that of a private fund. Brorsen and Irwin (1985) stated that the Securities and Exchange Commission (SEC) classifies a futures fund as public if limited partnership interests are sold in a public offering and as a rule of thumb, private funds have less than 35 investors. A private fund may be offered to only a select group of investors and is regulated as a private investment by the SEC unless specific tests are met. The CTA returns are from funds that are managed by individual CTA's, as opposed to the public and private funds which may have several CTA's. The CTA fund data was originally separated between those funds that are active (still being traded) and those funds that are dead (no longer being traded) but for this study the active and dead CTA data set were combined. The original data contained considerable missing values since many funds did not trade the entire period. These missing values were recorded as zeroes. The data source is LaPorte Asset Allocation. Most of the data originated from Managed Accounts Reports.

The data were cleaned of missing values by deleting observations where returns equaled zero and net asset value equaled zero and leverage equaled zero. The important note is that observations with a return of zero were not deleted because of the possibility that zero was the true return and should be included in the cleaned data set. Only data missing all values for all three zero conditions were deleted.

The relevant data used from the sets consisted of monthly percentage returns for different funds over time, the corresponding month and year of the return, and the name of the fund. Table 1 shows descriptive statistics of the data sets and of the variable of interest, the percentage return.

As with past research, the CTA funds have the highest mean returns. This may be a result of selectivity bias, but since this study is comparing CTAs to other CTAs and not to another type of investment, such as mutual funds, this is not a problem. The number of observations show that the CTA data set has the largest number of observations and number of funds, having over twice as many observations and funds as the private funds.

Table 1 shows that the CTA data has the highest variance of the fund types, followed by the private funds, with the public funds having the smallest variance. This can be interpreted to indicate that the CTA funds have the most risk associated with them. The results show that none of the fund types have skewness present but table 1 also shows that returns from all three fund types are leptokurtic.

Table 1. Descriptive Statistics for the Public, Private, and Combined CTA Data Sets and Continuous Time Returns

Data set	Public funds	Private funds	Combined CTA's
Observations	32420	23723	57018
# Funds	577	435	1071
Percentage returns			
Mean	0.31	0.62	1.28
SD	7.68	9.22	10.53
Minimum	-232.69	-224.81	-135.48
Maximum	229.73	188.93	239.79
Skewness	-2.08	-0.49	1.14
Kurtosis	133.91	40.70	24.34

Tests of Homoskedasticity and Normality

Two procedures were used to test the assumptions of homoskedasticity, normality, skewness, and relative kurtosis of the rescaled residuals. This procedure is necessary to determine if the rescaling of the residuals corrects the problem of heteroskedasticity. This study hypothesizes that the homoskedasticity assumption is incorrect. For this test, the null hypothesis is H_0: $Var(\epsilon_{i\,t}) = \sigma^2$ and the alternative is H_A : $Var(\epsilon_{i\,t}) = \sigma_i^2$.

The regression returns against fund dummy variables and an average monthly fund return using EGLS. The regression is specified as:

(4)

$$r_{jt} = \alpha_0 + \sum_{j=1}^{n-1} \alpha_j n_j + \sum_{j=1}^{n-1} n_j \beta_j \bar{r}_t + \epsilon_{jt} \qquad j = 1, \ldots, n-1$$

$$t = 1, \ldots, \textit{number of months}$$

where r_{jt} is the returns for fund j in month t, n_j is the fund dummy variable, and $\bar{r}_t$ is the monthly average return across funds.

Monthly average returns are calculated across funds. Returns are indexed by weighting each observation by the dollars invested in the fund. The value-weighted index is derived as:

(5)

$$\bar{r}_t = \frac{\sum_{i=1}^{nt} percret_{it} * Dol_i}{\sum_{i=1}^{nt} Dol_i}$$

where $\bar{r}_t$ is the percentage return in month t and Dol_i is the dollars under management of the fund.

The variance of the data is defined as:

(6) $$\epsilon_i^2 = \alpha_0 + \sum_{j=1}^{n-1} \alpha_j n_j$$

where n_j is a dummy variable for funds. There are n-1 dummy variables created in the regression, where n=the number of different funds, to represent the variation individual to that fund.

The regression method used was estimated generalized least squares (EGLS). The generalized least squares estimator used is:

(7) $$\hat{\beta} = (X'\Psi^{-1}X)^{-1}X'\Psi^{-1}y$$

where X is the matrix of fund dummy variables and monthly average fund return, y is the vector of returns, and Ψ represents the unknown covariance matrix that will be estimated. This method is necessary since heteroskedasticity is present.

Returns are regressed against a fund dummy variable and a month average return interaction variable (see equation 4). The residuals from this regression are squared and saved. These squared residuals are regressed against the fund dummy variables (see equation 6). From this regression, the predicted variances are calculated for the data. From this regression, an F-statistic is calculated for the hypothesis test that all the variances as a group are equal to zero. If the statistic shows significance, then the null of homoskedasticity is rejected. The procedure is the same for the public, private, and the combined CTA data sets.

Rescaling Residuals

To rescale the residuals, the reciprocal is taken for the predicted variances from the regression of the variance against the fund dummy variables (see equation 6):

$$(8) \qquad w_i = \frac{1}{\hat{\sigma}_i}$$

where the w_i are the square roots of the diagonal elements of the predicted variance. A possible problem area is the possibility of some funds having only one or two observations in the data set, therefore giving the fund zero variance and causing computational problems. To avoid having this problem, observations with a weight greater than 100 were deleted.

The final regression uses the model in equation 4. The percentage returns are now weighted by 1 over the predicted variance of the fund (see equation 8). Weighting allows for heteroskedasticity between the funds owing to the different risk levels of each fund.

Table 2 shows the results from the test for homoskedasticity. Results show the tests for the public, private, and CTA funds reject the H_0, indicating that heteroskedasticity is present in the data.

The next test is the normality test with the rescaled residuals. The residuals are from the regression of the returns on the fund dummy variables and fund dummies and average return interaction variables. These residuals were rescaled by dividing each residual by the square root of the predicted variance. From these rescaled residuals, the skewness and kurtosis are calculated for the rescaled residuals. The test statistic used is the Jarque-Bera test for normality. Under the null hypothesis that the error terms are distributed normally, the test statistic is distributed as a $\chi^2_{(2)}$. Clarke (1996) found that a normal distribution is not a correct assumption for managed futures funds, the returns exhibited higher skewness than returns for many derivative market indices. This was partially attributed to the nonstationary variance of the funds which would cause leptokurtosis.

Table 2 shows the results of the homoskedasticity and normality tests on the rescaled residuals. The null hypothesis of homoskedasticity was rejected for all three fund types. This confirms that heteroskedasticity is present in the residuals for each of the funds.

Table 2. F-Statistics for the Test of Homoskedasticity Assumption and Jarque-Bera Test of Normality of Rescaled Residuals

Data set	Public funds	Private funds	Combined CTA's
Homoskedasticity	1.41	4.32	5.15
Residuals			
Skewness	-0.17	-0.02	0.35
Relative Kurtosis	3.84	3.05	2.72
Normality of ϵ	20059*[a]	9192*	18735*

[a] Asterisks denote significance at the .005 level.

The results from the skewness tests show that none of the three fund types have skewed residuals. However, the tests show each fund type has kurtosis present in the residuals. The normality of the error terms is also rejected. Rescaled residuals actually have a t-distribution rather than a normal distribution. Several funds have less than 30 observations, thus, we should expect the remaining leptokurtosis which was found.

Testing If Fund Monthly Mean Returns as a Group Are Significantly Different from Zero

The first objective of this research is to answer a fundamental question regarding commodity funds, that being, are average monthly returns across funds the same after adjusting for returns on all funds? Heteroskedasticity is assumed to be present in all the data sets. Using regression techniques that are capable of handling heteroskedasticity by transforming the variables and correcting for this problem, an asymptotically valid test can be conducted.

The regression technique used was the same EGLS method used in the prior section. From the regression, a weighted ANOVA (analysis of variance) is completed. A type III analysis of variance is completed by using the SS3 option of PROC GLM command in SAS. A type III analysis of variance calculates the sum of squares for individual explanatory variables and calculates the sum of squares for interaction terms between explanatory variables. This is a regression with dummy variables, the dummy variables representing the individual mean returns from each different fund. A weighted ANOVA was performed to correct for the heteroskedasticity in the data. If the data are not weighted to reflect different variances, the heteroskedasticity present would cause the estimates to be inefficient and the hypothesis tests would be invalid. The regression first calculates the sum of squared errors for the individual means of the funds. These results show the percentage of total variation in returns that the individual fund mean returns explain. The sum of squared errors for monthly mean returns is then calculated. These results show the percentage of total variation in returns explained by the monthly mean return of all the funds.

In completing the ANOVA, computing a degrees of freedom adjustment for each fund for the test of normality is required. However, the degrees of freedom adjustment results in no gain in asymptotic properties.

A separate regression will be run on the public funds, the private funds, and the CTAs. The joint hypotheses to be tested from these estimates are: H_o: $\alpha_i = 0 \; \forall i$ and H_o: $\beta_i = 0 \; \forall i$.

Structure of Monte Carlo Study

Data Generation

Monte Carlo Study

The Monte Carlo simulation uses a data generating model based on the same model used to model the stochastic process that generates the returns. The data generating model is specified as:

(9) $$r_{it} = \alpha_i + \beta_i \bar{r}_t + \epsilon_{it} \quad i = 1, \ldots, n \qquad t = 1, \ldots, 120$$

$$\epsilon_{it} \sim N(0, \sigma_i^2).$$

Data sets were generated using α, β, and σ specified by the author. The α ,β, and σ matrices are t × 1 matrices where t = 120 different funds for which returns are being generated. From these values , 24 months of returns were generated for 120 funds. This procedure was repeated 1000 times. The mean return over all funds $\bar{r}_t$, is derived from the values of α and β. The equation for r_t is:

(10) $$\bar{r}_t = \frac{\frac{\sum \alpha_i}{n} + \frac{\sum \epsilon_{it}}{n}}{1 - \frac{\sum \beta_i}{n}} .$$

The data sets were generated using the Interactive Matrix Language (IML) module of SAS. The data sets were generated using a fixed value of α to simulate no performance persistence with a fixed seed of 31313 for the RANNOR command (random number generator) in SAS. The intercepts were set to 1 for the data generated with no persistence present. For the data sets generated with persistence present, α was allowed to vary and was generated using:

(11) $$\alpha_{CTA} \sim N(1.099,\ 4.99)$$

(12) $$\alpha_{Public} \sim N(.278,\ 1.35)$$

(13) $$\alpha_{Private} \sim N(.279,\ 5.20)$$

which were obtained from the GLM regressions (see equation 4) on the combined CTA data (equation 11), public funds (equation 12), and private (equation 13). These values differ from the values in table 1 since table 1 presents statistics using the raw returns.

Data were generated with different values for β. For the data generated with fixed β, a value of .5 was used. The differing β's included 4 different values. The β's took the new values in equal fourths. The four values of β are .5, 1, 1.5, and 2. These values encompass the range of most of the actual β's found.

Under homoskedasticity, data sets were generated with a fixed σ value of 2. Heteroskedasticity was imposed on the data sets by changing the value of σ from 2 to 4 different variances, these being 5, 10, 15, and 20. One fourth of the β's were set to each value. This allowed comparing the Spearman coefficient calculated for data sets with homoskedasticity and data sets with heteroskedasticity.

The funds were ranked in ascending order of returns for period one (first 12 months) and period two (last 12 months). From each 24 month period of generated returns, Spearman coefficients were calculated to identify the correlation between a fund's rank in period one and period two. For the Spearman coefficient, if the number of pairs of rankings is greater than ten, the distribution can be approximated by a normal distribution (Dowty and Wearden 1991). In this case, the number of funds that have a ranking in period one and period two are the number of pairs (120), therefore, the normal distribution is applicable. Spearman's coefficient (r_s) is calculated as:

(14) $$r_s = 1 - \frac{6\sum_{i=1}^{N} d_i^2}{N(N^2 - 1)}$$

where *d* is the difference in rank of a fund between period one and period two. The differences in rank are summed up across the total number of funds represented by N. The null hypothesis for the Spearman coefficient is:

(15) $$E(r_s) = 0$$

and the test statistic is calculated as:

(16) $$z = \frac{r_s - 0}{\sqrt{\frac{1}{N - 1}}} = r_s\sqrt{N - 1} \; .$$

From the calculated z values, the p-values were calculated for each repetition of the study. The p-value is the probability of obtaining a difference between the sample statistic and the hypothetical population parameter that is at least as extreme as the one actually observed assuming the H_0 is true. The smaller the p-value, the stronger the case against H_0. The p values were separated into three groups, those that failed to reject the H_o: $.025 < p \text{ value} < .975$, reject H_o with a positive z, $p \text{ value} < .025$ and those rejections with a negative z, $p \text{ value} > .975$. A percentage value was then calculated for all the repetitions using the formula:

$$(17) \qquad \frac{\textit{number of observations}}{\textit{number of repetitions of data generated}}$$

where the number of observations represents the number of observations from the generated data set that were statistically different from the observed value while the number of repetitions of data generated is the number of data sets generated by the simulation. These percentage values are being calculated to determine the size and power of the test. The power of the test is defined as 1 minus the probability of Type II error. Type II error is defined as failing to reject a H_0 when it is in fact false. The Monte Carlo simulations that generated data with persistence present by allowing the α's to vary refer to the power of a test. The size of a test refers to the probability of Type I error. Type I error is rejecting the H_0 when it is in fact true. The Monte Carlo simulations that generated data with no persistence present refer to the size of a test.

Mean returns were then calculated for each fund in period one and period two and ranked. The ranks were divided into three equal subgroups composed of the funds with the top third highest mean returns, middle third mean returns, and bottom third mean returns. Two additional subgroups were separated out, the top 3 highest mean returns funds and the bottom three funds with the lowest mean returns. The means across all funds in the top third group and bottom third group were calculated. A test of two means was done to determine if the returns from the top third funds and the bottom third funds are significantly different.

The power problems the EGR method is expected to have with means may not also exist in tests using Sharpe ratios. The Sharpe ratios include a measure of variance which allows for differing funds to have differing variances. The Monte Carlo simulation generated 24 and 60 months of monthly return data for 120 funds. Each sample was partitioned into selection period and performance period. The mean returns for each period were calculated. From the means, the sum of squares was calculated for each fund's returns during each period. The variance of each fund's returns was calculated. A simple Sharpe ratio was calculated for each fund by dividing the funds mean return by the standard deviation of the funds returns during that period. This process was repeated for both the selection period and the performance period. The Sharpe ratios were ranked in each period and the Spearman correlation coefficient was calculated between the selection period and the performance period. The Spearman correlation coefficient was tested for significance.

Historical Performance as an Indicator of Later Returns

4 Year Selection Period, 1 Year Performance Period

This procedure was to determine if any one of several historical performance measures provides information on the performance of a fund in subsequent years. The Monte Carlo study completed in this paper shows that the methods used by EGR have low power when evaluating the α's or mean returns when heteroskedasticity is present in the data. However, in light of this problem, this method is appropriate when the residuals are rescaled to compensate for the heteroskedasticity present in the data.

Gruber (1996) chose selection periods of 1 year and 3 years. This study chose a 4 year selection period with a 1 year performance period and was performed on the private, public, and CTA data sets. The 4 year selection period was chosen to allow several years of rankings to be included in the selection period as opposed to only one year to avoid the problem of short-run negative autocorrelation. The same procedure to correct for heteroskedasticity in objective one is also used in determining if historical performance is an indicator of future returns. Since the returns are monthly, funds that had fewer than 60 monthly observations were deleted to avoid having missing months of data.

A generalized linear model (see equation 4) was used to regress the returns against the average fund return. The data were split into the 4 year selection period and the 1 year performance period. The first 5 year period evaluated was 1980-84. The periods were repeated until the final period, 1991-1995. Testing for persistence was done on three parameters derived from the regression, the α 's (intercept), the mean returns, and α / σ (adjusted Sharpe Ratio). For each parameter derived from the regression, a Spearman coefficient was calculated between the rank of the parameter from the 4 year selection period and the rank of the parameter for the 1 year performance period. The same coefficient was calculated for the mean return and the α / σ. The Spearman coefficient, (see equation 14) was calculated. The null hypothesis of an expected value of 0 for the Spearman correlation coefficient (see equation 15) can be tested with a z test (see equation 16). The z-statistic was calculated and evaluated at α=.05.

Historical Performance as an Indicator of Later Returns

Two-Way Tables of Multiple Selection and Performance Periods

The second method of evaluation used the method of Goetzmann and Ibbotson (1994). This method used raw return data from predetermined selection and performance periods. The method compared the classification of the mean fund return as a winner or loser in each period. Data were separated into chronologically nonoverlapping samples which included selection period and performance period. Each sample was then separated into returns from the selection period and returns in the performance period.

The data were then sorted by fund name. Each fund then had the simple mean calculated for its returns in each period. The simple mean was calculated by summing the monthly returns across the selection period and dividing by the number of monthly returns in the selection period. This is referred to from this point on as the selection mean return. For the performance period, the monthly returns across the performance period were summed and divided by the number of monthly returns in the performance period. This is referred to as the performance mean return. Each fund now has a selection mean return and a performance mean return.

The median is calculated for the individual fund selection mean returns. From this median mean fund return, funds are classified as winners or losers. The classification as a winner is mean return of the fund is greater than the median. The classification as a loser is mean return of the fund is less than the median. The median is then calculated for the individual fund performance mean returns. The classification as a winner or loser is based on the median of the performance period. The use of the median as the criteria for classification as winner/ loser is based on the methods used by Goetzmann and Ibbotson.

The results are then reported in two-way tables of winners and losers for each period. Results show funds are classified into one of four subgroups: winner/ winner, loser /loser, loser/ winner, and winner/ loser. The winner/ winner group reflects the number of funds classified as a winner in the selection period and also classified as a winner in the performance period. The loser/ loser group shows the number of funds classified as losers in the selection period and classified as loser in the performance period. The final two groups, loser/ winner and winner/ loser reflect the number of funds that respectively, fell into each classification based on performance.

A two-way table was chosen for the simplicity and ease of interpretation of the results. Cumulative numbers of funds that fell into each category are presented at the end of each two-way table. The percentages in the cumulative results represent the percent of funds in each performance period classification (winner/ loser) as a percentage of the funds in each selection period classification. For winners in the selection period, what percent of winners also won in the performance period and what percentage lost. From these percentages, a ratio of the likelihood of a fund continuing prior performance is calculated.

For robustness of the results, selection periods and performance periods were varied. The periods tested included consecutive two-year periods. The data was partitioned into four-year nonoverlapping samples. Each four year block was then partitioned into two two-year periods. The chronologically earlier two-year period was designated the selection period. The later two-year period was designated the performance period. The procedure was repeated for consecutive three-year fund returns and consecutive one-year fund returns. The procedure then chose a three-year selection period and two-year performance period, and two-year selection periods with a one-year performance period. In each variation of the length of the selection and performance period, if, at the end of the data set, the remaining years of data did not form a complete sample (three-year selection period and three-year performance period), then the remaining data was deleted.

The hypothesis of no persistence is that the ratio of winners in the selection period to winners in the performance period should be 50/50. A fund that is a winner in the selection period is just as likely to be a loser in the performance period. This hypothesis is tested by evaluating the ratios of winners in the selection period to winners in the performance period, with the same ratio calculated for losers in the selection period who are also losers in the performance period.

CHAPTER IV

RESULTS

Joint Test of No Difference in α's across Funds

Table 3 shows a weighted ANOVA table of the general linear model. Table 3 also shows the mean and variance of α which are used in the Monte Carlo study. It also shows the F-statistic values calculated for the joint tests of no difference in α's across funds and the values for the joint tests of no difference in β's across funds. The results show that funds and pools do not all have the same mean returns. This finding is not consistent with prior research, but not surprising for several reasons. Following the efficient market hypothesis there may be several reasons why all funds do not have the same mean returns.

Table 3. Weighted ANOVA Table: Returns Regression for Public Funds, Private Funds, and Combined CTA Data

Fund Type	Public funds	Private funds	Combined CTA's
Sum of Squared Errors			
Ind. means	1751	1948	2333
Group mean	28335	10882	22751
Corrected Total	62221	36375	82408
R^2	0.48	0.35	0.31
Mean			
α	0.278	0.297	1.099
Variance			
α	1.16	2.277	2.240
F-statistic			
α's	2.94	4.32	2.12
β's	47.44	24.10	20.61

Efficient markets would dictate that superior performing funds would have capital flow into the fund from poorer performing funds assuming no transaction costs. This shift of capital would have two effects. The funds with capital flowing into them would have their returns decrease and the funds with capital flowing out of them would see their returns increase. However, there are transaction costs to shifting capital from one fund to another. Load fees, withdrawal fees, and penalties for shifting capital may make it costly to move the capital to the better performing fund. Investors may also choose to shift the capital out of the managed futures market, in which case capital still does not flow from poorer performing funds to better performing funds.

Another reason for capital not shifting from poorer performing funds to better performing funds is that the better performing fund may be closed. Closed meaning that the fund is not accepting any new investments. Capital from poorer performing funds could flow into the better performing fund, but since no new investment is being taken, some capital from poorer performing funds must go somewhere else.

The results show that the individual means do explain some of the variance of the returns. This supports the contention that some funds earn significantly different returns than other funds do across years. The differences in individual means explain roughly 2-4% of the variance in the public funds. This is a small amount of predictability and requires precise methods to take advantage of it.

Analysis of Power of EGR Method

A Monte Carlo analysis of the power and size of EGR's method is now presented. Table 4 shows the calculated mean returns from each subgroup of funds generated with no persistence present. The results indicate the size of the test when no heteroskedasticity is present. There is a slight tendency to reject too often when heteroskedasticity is present. The results show the size is slightly low but still reasonable in the test of two means.

Table 4. Performance Persistence Results (EGR method) for Mean Returns Over Consecutive 1-Year Periods from Monte Carlo Generated Data Sets: No Persistence Present by Fixing α's to 1

	Data Generation method		
Generated Data Subgroups	1[a]	2[b]	3[c]
Mean returns			
top 1/3	1.25	1.25	0.70
middle 1/3	1.25	1.25	0.72
bottom 1/3	1.25	1.22	0.68
top 3	1.25	1.15	0.61
bottom 3	1.26	1.19	0.68
p-values			
reject-positive z	.021	.041	.041
reject-negative z	.028	.037	.039
fail to reject	.951	.922	.920
test of 2 means			
reject-positive	.026	.032	.032
reject-negative	.028	.020	.026
fail to reject	.946	.948	.942

[a]Data generated using $\alpha=1$, $\beta=.5$, $\sigma=2$.
[b]Data generated using $\alpha=1$, $\beta=.5$, $\sigma=5$, 10, 15, 20.
[c]Data generated using $\alpha=1$, $\beta=.5$, 1, 1.5, 2, $\sigma=5$, 10, 15, 20.

Table 5 shows the calculated Sharpe ratios over consecutive one-year periods generated with no persistence present. The results again indicate the size of the test is correct when no heteroskedasticity is present. However, when heteroskedasticity is present, the null hypothesis is rejected more than 5% of the time. Most of the rejections favor positive correlations. Thus, with Sharpe ratios, rejection of performance persistence can occur when the performance persistence is in the variance, but not the mean. The results show the size is slightly low but still reasonable in the test of two means.

Table 5. Performance Persistence Results (EGR method) for Sharpe Ratio Over Consecutive 1-Year Periods from Monte Carlo Generated Data Sets: No Persistence in Mean Present by Fixing α's to 1

	Data Generation method		
Generated Data Subgroups	1[a]	2[b]	3[c]
Mean returns			
top 1/3	0.67	0.15	0.09
middle 1/3	0.66	0.14	0.08
bottom 1/3	0.66	0.13	0.08
top 3	0.66	0.16	0.10
bottom 3	0.66	0.11	0.07
p-values			
reject-positive z	.024	.095	.055
reject-negative z	.023	.004	.009
fail to reject	.953	.901	.936
test of 2 means			
reject-positive	.029	.046	.042
reject-negative	.027	.013	.016
fail to reject	.944	.941	.942

[a]Data generated using α=1, β=.5, σ=2.
[b]Data generated using α=1, β=.5, σ=5, 10, 15, 20.
[c]Data generated using α=1, β=.5, 1, 1.5, 2, σ=5, 10, 15, 20.

Table 6 shows the calculated Sharpe ratios over a four- year selection period and a one-year performance period from each subgroup of funds generated with no persistence present. As with the prior results, the size of the test is acceptable when no heteroskedasticity is present. When heteroskedasticity is present, the tests reject the null hypothesis more often than with the one-year performance period. As before, the results show the size is slightly low but still reasonable in the test of two means.

Table 6. Performance Persistence Results (EGR method) for Sharpe Ratio Over Four-Year Selection Periods and 1-Year Performance Periods from Monte Carlo Generated Data Sets: No Persistence Present in Mean by Fixing α's to 1

	Data Generation method		
Generated Data Subgroups	1[a]	2[b]	3[c]
Mean returns			
top 1/3	0.67	0.16	0.10
middle 1/3	0.66	0.14	0.08
bottom 1/3	0.67	0.12	0.07
top 3	0.67	0.18	0.12
bottom 3	0.67	0.11	0.06
p-values			
reject-positive z	.025	.206	.115
reject-negative z	.024	.001	.004
fail to reject	.951	.793	.881
test of 2 means			
reject-positive	.024	.071	.047
reject-negative	.030	.008	.008
fail to reject	.946	.921	.945

[a]Data generated using $\alpha=1$, $\beta=.5$, $\sigma=2$.
[b]Data generated using $\alpha=1$, $\beta=.5$, $\sigma=5$, 10, 15, 20.
[c]Data generated using $\alpha=1$, $\beta=.5$, 1, 1.5, 2, $\sigma=5$, 10, 15, 20.

Table 7 shows the results of testing EGR's ability to find performance persistence when it really exists in mean returns over consecutive one-year periods. Although the results differ somewhat across all columns, the results show that overall, the power of the Spearman's coefficient is high in all columns except for column four. Column one shows that the null hypothesis is always rejected. Columns two and three show the null hypothesis rejected over 80% of the time. Only column four shows a substantial failure

(.848) to reject the null where heteroskedasticity is present and the performance

Table 7. Performance Persistence Results (EGR method) for Mean Returns Over Consecutive 1-Year Periods from Monte Carlo Generated Data Sets: Persistence Present by Allowing α's to Vary

	Data Generation method			
Generated Data Subgroups	1[a]	2[b]	3[c]	4[d]
Mean returns				
top 1/3	3.21	2.77	2.57	1.48
middle 1/3	1.87	2.09	1.85	1.30
bottom 1/3	0.80	1.41	1.15	1.14
top 3	4.93	3.47	3.26	1.68
bottom 3	-1.60	1.14	0.86	1.06
p-values				
reject-positive z	1.00	.827	.823	.149
reject-negative z	.000	.000	.000	.003
fail to reject	.000	.173	.177	.848
test of 2 means				
reject-positive	1.00	.268	.258	.043
reject-negative	.000	.000	.000	.012
fail to reject	.000	.732	.742	.945

[a]Data generated using α=N(1.099,4.99), β=.5, 1, 1.5, 2, σ=2.
[b]Data generated using α=N(1.099,4.99), β=.5, σ=5, 10, 15, 20.
[c]Data generated using α=N(1.099,4.99), β=.5, 1, 1.5, 2, σ=5, 10, 15, 20.
[d]Data generated using α=N(1.099,1), β=.5, 1, 1.5, 2, σ=5, 10, 15, 20.

persistence is small.

The results from the test of two means show that the test works poorly with heteroskedasticity present. This shows EGR's nonparametric method of using Spearman correlation coefficients on performance measure variables with no adjustment for heteroskedasticity does have substantial power problems and is inappropriate for testing for performance persistence. Table 8 shows the results of testing EGR's ability to find performance persistence when it really exists in Sharpe ratios over consecutive one-year

Table 8. Performance Persistence Results (EGR method) for Sharpe Ratio Over Consecutive 1-Year Periods from Monte Carlo Generated Data Sets: Persistence Present by Allowing α's to Vary

	Data Generation method			
Generated Data Subgroups	1[a]	2[b]	3[c]	4[d]
Mean returns				
top 1/3	1.66	0.31	0.27	0.16
middle 1/3	1.01	0.21	0.18	0.13
bottom 1/3	0.42	0.13	0.09	0.11
top 3	2.36	0.46	0.42	0.21
bottom 3	-0.65	-0.04	-0.09	0.08
p-values				
reject-positive z	1.00	.995	.993	.349
reject-negative z	.000	.000	.000	.000
fail to reject	.000	.005	.007	.651
test of 2 means				
reject-positive	1.00	.494	.471	.114
reject-negative	.000	.000	.000	.001
fail to reject	.000	.506	.529	.885

[a]Data generated using α=N(1.099,4.99), β=.5, 1, 1.5, 2, σ=2.
[b]Data generated using α=N(1.099,4.99), β=.5, σ=5, 10, 15, 20.
[c]Data generated using α=N(1.099,4.99), β=.5, 1, 1.5, 2, σ=5, 10, 15, 20.
[d]Data generated using α=N(1.099,1), β=.5, 1, 1.5, 2, σ=5, 10, 15, 20.

periods. As with table 7, results differ across the columns, but the power of the Spearman's coefficient is high in all columns except for column four. Column four again showed a substantial failure (.651) to reject the null. Thus, even EGR's tests with Sharpe ratios had little power. Also, the test of two means works poorly with heteroskedasticity present. Table 9 shows the results of testing EGR's ability to find performance persistence when it really exists in Sharpe ratios over a four-year selection period and a one-year performance period. The results differ from tables 7 and 8 in that the power of the

Table 9. Performance Persistence Results (EGR method) for Sharpe Ratio Over Four-Year Selection Periods and 1-Year Performance Periods from Monte Carlo Generated Data Sets: Persistence Present by Allowing α's to Vary

	Data Generation method			
Generated Data Subgroups	1[a]	2[b]	3[c]	4[d]
Mean returns				
top 1/3	1.71	0.34	0.30	0.18
middle 1/3	0.98	0.21	0.18	0.12
bottom 1/3	0.41	0.11	0.07	0.09
top 3	2.55	0.56	0.49	0.26
bottom 3	-0.91	-0.15	-0.22	0.04
p-values				
reject-positive z	1.00	1.00	1.00	.711
reject-negative z	.000	.000	.000	.000
fail to reject	.000	.000	.000	.289
test of 2 means				
reject-positive	1.00	.712	.695	.191
reject-negative	.000	.000	.000	.000
fail to reject	.000	.288	.305	.809

[a]Data generated using α=N(1.099,4.99), β=.5, 1, 1.5, 2, σ=2.
[b]Data generated using α=N(1.099,4.99), β=.5, σ=5, 10, 15, 20.
[c]Data generated using α=N(1.099,4.99), β=.5, 1, 1.5, 2, σ=5, 10, 15, 20.
[d]Data generated using α=N(1.099,1), β=.5, 1, 1.5, 2, σ=5, 10, 15, 20.

Spearman coefficient is good across all four columns. This supports our use of the four-year selection and one-year performance periods. However, the test of two means shows low power.

Historical Performance as a Predictor of Future Performance

Table 10 shows a summary of the results from the out-of-sample testing. Appendix Tables 1- 9 show the results for each year. Spearman correlation coefficients calculated between the selection period (4 years) and the performance period (1 year). The results show the average correlation calculated between the variable used in the selection period and the same variable from the performance period. The years positive show the percentage of years where the correlation was positive. A positive correlation indicates performance persistence was present. This follows that a fund that did well during the selection period also does well during the performance period. Negative correlations reflect the opposite of performance persistence, a fund that does well during the selection period does poorly during the performance period. The final column shows the percentage of years where the correlation was positive and significant.

Results show each fund type contains a high percentage of years with positive correlations between the selection period and performance period. CTA funds show mean returns and the α's being positively correlated over 83% of the sample. The adjusted Sharpe ratio shows positive correlation between the selection period and the performance period in every sample.

Table 10 shows the percentage of years with correlation positive and statistically significant. With the CTA data, rankings by both mean returns and α's were positively correlated and significant 25% of the time. The adjusted Sharpe ratios showed a higher percentage of years with positive correlation that were significant at 42%. The high percentages of positive correlations are what was expected, a positive correlation between

Table 10. Summary of Spearman Correlations between Selection and Performance Periods

Data set selection criterion	Average correlation	Years positive (%)	Years Positive and significant (%)
CTA			
mean returns	0.118[a]	83	25
α	0.114	83	25
α/σ	0.168	100	42
Public funds			
mean returns	0.084	75	33
α	0.088	75	33
α/σ	0.202	83	42
Private funds			
mean returns	0.068	58	17
α	0.047	58	0
α/σ	0.322	92	50

[a]Correlation between a four-year selection period and a one-year performance period. Averages are across the twelve one-year performance period. The same statistic was used for the rankings in each period.

the measurement variables in the selection period and the performance period. This confirms the previous results of a small amount of performance persistence.

The average correlations of the performance measures show that for the CTA funds, the mean returns and the α's from the selection period explain slightly more than 10% of the variation of the performance period for mean returns and α's. While not perfect positive correlation, this amount of correlation could still be useful in helping to select a fund in which to invest. The Sharpe ratio explains a larger amount, almost 17% of the variation. Results from the public funds show that the public funds do not have as high a percentage of positive correlations across the performance measures tested as the CTA funds.

The average correlation of the public fund mean returns and α's is smaller than that of the CTA funds. The mean returns showed 8% and the α's showed 9% of the variation explained. These amounts could still be used in assisting in the selection of a fund. However, the adjusted Sharpe ratio has an average correlation of approximately 20%.

Results from the private funds are similar. The mean returns and the α's have lower percentages of positive correlations than the CTA and public funds. Both measures had 58% of the years with positive correlations. As with the CTA data and the public funds, the adjusted Sharpe ratio had the highest percentage of years with positive correlations at 92%.

For the private funds, the mean returns showed only 17% of the years with positive correlations as being significant. The α's for private funds showed no significant positive correlations. However, the adjusted Sharpe ratio showed that 50% of the years with positive correlations were significant. This appears to indicate that the Sharpe ratio exhibits the most persistence of any of the three performance measures tested.

Results from the private funds show that the mean returns and the α's have the lowest average correlation of any of the three funds. Mean returns have only 7% of the variation explained while the α's has only 5% explained. The adjusted Sharpe ratios have the highest average correlation of any variable in any fund with over 32% of the variation explained. This could be a valuable tool in predicting future Sharpe ratios for funds with past Sharpe ratio values.

Results indicate that in CTA funds, private funds, and public funds, mean returns and α's can be indicators of future performance, but with limited success. These two variables show that some persistence is present in regards to future performance. However, it should be noted that the adjusted Sharpe ratios for each of the funds seemed to exhibit the most persistence of any of the performance variables. For all the fund types, the measure of risk variable (adjusted Sharpe's ratio) is a good indicator of future Sharpe ratios. Past research has found risk measures often had persistence present. Positive correlation indicates risk measures are a good measure of a funds risk in trading.

Regardless of the performance measure used, there is some positive correlation indicating performance persistence. The small correlations are consistent with the regression results. While there is performance persistence present, it is difficult to distinguish because of all the other random factors involved in influencing returns.

The return/risk measure (adjusted Sharpe Ratio) clearly shows the most performance persistence. Rankings based on mean returns are similar to the rankings based on α's. As the appendix shows, their correlations are similar in each year. Therefore, there does not appear to be as much gain as expected in adjusting for the overall level of returns.

Appendix Tables 1 through 9 show the individual selection and performance period results. Appendix Tables 1-3 show the results from the CTA data. Mean returns and α's show negative correlations between selection period and the performance period for two periods, but neither is significant. Adjusted Sharpe ratio results show no negative correlations in any of the periods. All three measures have significant positive correlations in the same three periods. The strongest persistence appears in the adjusted Sharpe ratio.

Appendix Tables 4-6 show results from the public fund data. The α's show three periods with negative correlation, one of which was significant. However, four periods were found to have significant positive correlations, all of which are greater than .25. Results from the mean returns correlation show four periods with significant positive correlations with values all greater than .25. Mean returns and α's both exhibit small amounts of persistence. The adjusted Sharpe ratio had five periods with significant positive correlations, more than either of the other performance measures, indicating more persistence was present..

Appendix Tables 7-9 show the results from the private fund data. Results from the α's showed periods with negative and positive correlations, none of which were significant. Returns showed two periods with significant positive correlations. The adjusted Sharpe ratios exhibited the most persistence.

This brings up the question as to why these results differ from past results? EGR dismissed evidence of persistence as small and insignificant. However, here the percentage of years with a statistically significant positive correlation is too large to dismiss. While the size of the correlations may be small relative to a correlation of one, the correlations are larger than expected if there were no performance persistence. Also, the adjusted Sharpe ratios have the largest average correlations of any measure. This correlation indicates that some persistence exists in the risk levels for the funds. Larger samples now available now allow more powerful tests.

McCarthy found persistence present but discounted the results due to the small sample size. IKZ used a method which separated the funds into quintiles. This method led to low power and difficulty in interpretation and therefore it is difficult to say whether a positive or negative correlation was found. Schwager found a similar correlation of .07 for mean returns. Schwager, however, found a negative correlation for his return/risk measure. Schwager ranked funds based on return/risk when returns were positive, but ranked on returns only when returns were negative. This hybrid measure may have led to the negative correlation. Therefore, past literature is consistent with a minimal amount of performance persistence. The large sample size and improved testing methods allowed us to find performance persistence. The appearance of negative correlations between selection periods and performance periods coupled with insignificant positive correlations show that small samples yield erratic results.

Historical Performance as a Predictor of Future Performance

Goetzmann and Ibbotson Nonparametric Method

Tables 11-15 present the results from evaluating if historical mean returns are predictive of future returns using the methods used by Goetzmann and Ibbotson. The results are presented in the form of two-way tables. The two-way tables present the number of funds in each classification, winner or loser in the selection period and winner or loser in the performance period. Table 11 shows the results from the two-year selection and performance period using the mean raw returns as a performance measure. Table 11 shows funds that were classified as winners in the selection period were more likely to also be winners in the performance period. The cumulative results show the combined results of all periods, indicating the ratio associated with picking a winner. The percentages presented below the summed totals represent the percentage of total funds that fall into

Table 11. Two-Way Tables of Ranked CTA Fund Raw Returns over Successive Two-Year Intervals

	1982-1983 Winners	1982-1983 Losers		1986-1987 Winners	1986-1987 Losers
1980-1981			1984-1985		
Winners	13	6	Winners	25	14
Losers	6	10	Losers	14	25

	1990-1991 Winners	1990-1991 Losers		1994-1995 Winners	1994-1995 Losers
1988-1989			1992-1993		
Winners	44	37	Winners	71	47
Losers	37	44	Losers	47	71

	Combined Results Over All Years Two Year Successive Periods Winners	Losers
Selection Period		
Winners	153 59.5%	104 40.5%
Losers	104 40.9%	150 59.1%

each selection period and subsequent performance period classification (winner or loser). The winners-winners category shows that winners in the selection period were also winners in the performance period roughly 60% of the time.

Table 12 shows the results from the three-year selection and performance period using the raw returns as a performance measure. The cumulative results show that like the successive two-year intervals, winners in the selection period are likely to be winners in the performance period roughly 60% of the time. The percentages of winners in the selection period that also win in the performance period are consistent across the samples. The ratios are roughly 60/40 in both samples, confirming that persistence is present.

Table 12. Two-Way Tables of Ranked CTA Fund Raw Returns over Successive Three-Year Intervals

	1983-1985 Winners	1983-1985 Losers		1989-1991 Winners	1989-1991 Losers
1980-1982			1986-1988		
Winners	9	5	Winners	32	22
Losers	6	9	Losers	22	33

	Combined Results Over All Years Three Year Successive Periods Winners	Losers
Selection Period		
Winners	41 60.3%	27 39.7%
Losers	28 40%	42 60%

Table 13 shows the results from the three-year selection period and two-year performance period using the raw returns as a performance measure. The results from the 1980-1984 sample may be split to a nearly 50/50 ratio due to the small number of funds in

Table 13. Two-Way Tables of Ranked CTA Fund Raw Returns over a Three Year Selection Period and a Two-Year Performance Period

	1983-1984 Winners	1983-1984 Losers		1988-1989 Winners	1988-1989 Losers
1980-1982			1985-1987		
Winners	8	7	Winners	30	21
Losers	8	8	Losers	21	30
	1993-1994 Winners	**1993-1994 Losers**		**Combined Results All Periods Winners**	**Losers**
1990-1992			Selection Period		
Winners	70	45	Winners	108 59.7%	73 40.3%
Losers	45	70	Losers	74 40.7%	108 59.3%

the sample. Table 13 confirms what the prior two tables found, winners in the selection period are winners in the performance period roughly 60% of the time.

When selection and performance periods are set to contain multiple years, some persistence of performance is present. The results show for each of the selection and performance scenarios, a fund's relative performance in the selection period has some predictive ability as to the relative performance of that fund in the future. Winners tend to continue as winners more often than losing.

Table 14 shows the results from the one year selection and performance period using the raw returns as a performance measure. The results from this table show that winners in the selection period have roughly a 50% chance of also being a winner in the performance period. In three of the eight samples, winners in the selection period exhibit a greater chance of being winners (≈ 60%)in the performance period. However, four of the samples showed roughly a 50% chance of being winners in the performance period after being classified a winner in the selection period. The final sample showed a 45% chance of winners repeating. Overall, there is no consistent persistence found. Based on the cumulative results of a one-year selection period and a one-year performance period, fund classification in the selection period has little predictive ability as to the fund's performance in the future. The results indicate that a funds future performance could not be accurately predicted from past performance. These results are consistent with the hypothesis of no performance persistence.

Table 15 shows the results from the two year selection period and a one year performance period using the raw returns as a performance measure. The results are consistent with the results from Table 14. Winners in the selection period have a roughly

Table 14. Two-Way Tables of Ranked CTA Fund Raw Returns over a One-Year Selection Period and a One-Year Performance Period

		1981 Winners	1981 Losers			1983 Winners	1983 Losers
1980				1982			
	Winners	10	10		Winners	18	12
	Losers	10	11		Losers	11	18
		1985 Winners	1985 Losers			1987 Winners	Losers
1984				1986			
	Winners	23	20		Winners	41	39
	Losers	21	23		Losers	39	40
		1989 Winners	1989 Losers			1991 Winners	1991 Losers
1988				1990			
	Winners	58	44		Winners	81	55
	Losers	44	58		Losers	55	81
		1993 Winners	1993 Losers			1995 Winners	1995 Losers
1992				1994			
	Winners	90	92		Winners	83	103
	Losers	92	90		Losers	102	83

	Combined Results Over All Years One Year Successive Periods Winners	 Losers
Selection Period		
Winners	404 51.9%	375 48.1%
Losers	374 48.1%	404 51.9%

50% chance of also being winners in the performance period. It is noted that the two most recent samples seem to exhibit some persistence with winners having roughly a 60% chance of being a winner in the performance period. The early samples show little or no signs of persistence, however.

The results show persistence is present in mean returns when multiple year selection and performance periods are selected. Persistence is not present when the performance period is set to one year. The results do not change as the selection period is extended from one to four years, therefore demonstrating the results are robust.

Table 15. Two-Way Tables of Ranked CTA Fund Raw Returns over a Two-Year Selection Period and a One-Year Performance Period

	1982 Winners	1982 Losers		1985 Winners	1985 Losers
1980-1981			1983-1984		
Winners	9	11	Winners	10	18
Losers	10	9	Losers	20	11
	1988 Winners	**1988 Losers**		**1991 Winners**	**1991 Losers**
1986-1987			1989-1990		
Winners	38	36	Winners	55	46
Losers	36	39	Losers	47	55
	1994 Winners	**1994 Losers**		**Combined Results All Periods Winners**	**Losers**
1992-1993			Selection Period		
Winners	58	44	Winners	193 51.3%	183 48.7%
Losers	44	58	Losers	186 48.8%	195 51.2%

CHAPTER V

SUMMARY

Commodity fund managers are required by law to release records of past performance in the prospectus of the funds they manage. This raises the question, is past performance indicative of future performance? Can past measures of performance predict how a fund will do in the future? This study analyzed a method used by prior research to determine if a key assumption of homoskedasticity was correct. Using a selection period and a performance period, several measures of performance were tested to determine if past fund performance is correlated with future fund performance.

One of EGR's methods assumes homoskedasticity and data reported disputes the assumption of homoskedasticity in fund returns. By using estimated generalized least squares (EGLS) to correct for heteroskedasticity, the problem is alleviated. When the problem is corrected, the EGLS regression analysis shows that all fund returns do not have the same mean.

The Monte Carlo simulation addressed the question of whether EGR's method of testing for performance persistence has any power to reject a false null hypothesis of no performance persistence in mean returns. EGR's method consisted of a rank correlation coefficient. The test statistic calculated was the Spearman correlation coefficient. The test statistic was calculated by ranking the variable of interest during the selection period and the performance period, then calculating the correlation between the rankings. The amount of correlation between the measure of performance in the selection period and the measure of performance in the performance period and the statistical significance of the correlation coefficients were then evaluated.

Data sets were generated with persistence present and with no persistence present. The Monte Carlo study also generated and tested several data sets where assumptions made by EGR's method are violated. The test revealed the method used by EGR has high power when persistence is present and heteroskedasticity is not. The Monte Carlo study showed that heteroskedasticity created some problems with size, but the problems were small. EGR's methods had good power except when the performance persistence was small and heteroskedasticity was present (which was the case with EGR's data). EGR's method has greater power when the Sharpe ratio was the measure of performance and the selection period was four years and the performance period was one year. The weakness of the Sharpe ratio is that persistence can be due either to persistence in mean or persistence in variance. The assumption of homoskedasticity made in the EGR study may have caused no persistence to be found where persistence was present. This could lead to the incorrect conclusion by investors that performance persistence does not exist in mean returns or the α's.

Three measures of historical performance were used to select which funds performed the best during the selection period. The measures chosen for evaluation of performance persistence are the returns of the fund, the α's (intercept) for each fund, and an adjusted Sharpe ratio (α/σ) to measure the risk of the fund. These measures had the rank correlation calculated between the value of that variable during the selection period and the performance period. The Spearman coefficient was calculated and it was found that the return/risk measure showed the most predictability in ranking from the selection period to the performance period.

Even with the heteroskedasticity corrected, while the results do not conclusively show that the intercepts and the mean returns of the funds are good indicators of later performance of funds, the results do suggest that the intercepts and mean returns have some predictive ability. The information contained in these variables appears to have a small but valid ability to help in identifying superior performing funds. The α / σ's consistently show that they do have some predictive ability. The inclusion of the variability of the fund seems to add relevant information that does have some predictive value. This measure does show that by including returns and variability, this can help an investor select the better performing funds.

It should be pointed out that the regression approach and not the rank correlations used in this study, allows funds that have been traded for differing amounts of time to be compared and allows for use of all the data. This differs from past methods in that those methods can only use funds that have been traded during the same amount of time.

Using Goetzmann and Ibbotsons method on CTA fund data, results showed that winners tend to stay winners by a ratio of roughly 60/40 when the selection period and performance period are both greater than one year. This is consistent with the ratio found in mutual funds. The ratio of winners staying winners fell to 50/50 when a one-year performance period was chosen. This ratio was relatively unchanged for a selection period of one, two, three, or four years. The results appear to show that funds with mean returns above the median fund return for the selection period tend to have mean returns that are above the median fund return for the performance period. This persistence is present when multiple-year selection periods and multiple-year performance periods are evaluated. The persistence is not present when one-year performance periods are selected, regardless of the length of the selection period.

Limitations of the Study

This study chose to use a four year selection period in testing for performance persistence using EGR's methods. This selection period is a narrowly defined period of performance evaluation and does not imply that different selection periods, longer or shorter, could not be useful in finding the optimal selection period.

There may still be a small problem with survivorship bias. The survivorship bias is caused by a fund no longer being traded. This could be the result of several things. The fund may have consistently performed poorly and lost most or all of the capital invested in it and went out of business. The fund may also have been bought out by another fund or merged with another fund. In both cases, the bought-out fund is not available anymore to investors and its returns are not known.

This survivorship bias comes in the form of testing for performance persistence will be conducted only on the "surviving" funds. The "surviving" funds are those funds that either have not been bought up or gone out of business. Performance persistence may exist in the funds that did poorly, however, since the fund is no longer traded, the persistence may not be found. For funds that are bought up by other funds, no distinction is made between bought-out funds and funds that went out of business, the fund simply stops trading. The funds with poorer track records tend to disappear over time. The fund may have been performing well and hence, was purchased by another fund and merged. The merged fund may have had persistence present, but it may not be found since the fund was bought and merged with another fund.

However, since performance is not measured against an absolute benchmark such as a market-based index, the S&P 500, but rather against the other "survivor" funds for relative performance, the problem may be partially mitigated.

Suggestions for Further Research

For the selection period, a more diverse variety and evaluation of selection periods could be completed including shorter selection periods and longer selection periods. It may also interesting to evaluate if the performance period has a lag greater than one year from the selection period. This is to say that the selection period predicts performance in the latter part of the performance period. This would be interesting to find if changes in the fund take time to develop and show positive returns. This asks the question, are changes made in the fund instantaneously reflected in higher returns?

Another line of research that may be studied from this area is to study the length of time a fund that consistently does poorly lasts before it is dropped. This area could be helpful in determining the plateau investors have in fund returns that gives the investor incentive to withdraw from a fund that is doing poorly. Ideally, the investor tracks their fund closely, but after how much time with poor returns, does the investor decide to withdraw from the fund and invest elsewhere? Is there a timetable that the investor has in mind as to how long the investor will give the fund manager to earn returns that are considered satisfactory?

Conclusions

The primary conclusion is that some funds consistently have higher returns than others. While the differences are small, the differences are statistically significant and were found with a variety of methods.

The second point is that past studies have made assumptions about the data that are not correct for evaluating performance variables that do not include some measure of risk. By making these assumptions, the prior studies results are not strictly valid for performance variables with no risk measure included. Using a nonparametric method in evaluating persistence is acceptable when the correct assumptions are made, the variables are transformed to remove the problem of an incorrect assumption, or the performance variables includes a measure of risk.

The third point is that after correcting the data for heteroskedasticity, the study shows that the Sharpe ratio does have some predictive ability as to the future returns of the fund. This shows some of the past research results using Sharpe ratios are valid. The predictive ability is shown to be present in the private, public, and combined CTA data. Data indicate that investors would want to look at historical Sharpe ratios as a factor in determining in which funds to invest. For this study, the prior 4 years Sharpe ratios had significant predictive ability, but the data suggest that the data from the last year should be deleted. If managing money, the rationale should be to include the prior 4 years Sharpe ratios in the decision making process because this measure has shown to have some predictive ability. However, it should be noted that longer selection periods should not be dismissed as useless. Although not perfect predictors, by explaining even 2% of the variation of the mean return investor return can be improved over time.

The ability to explain a small amount of the variation in fund returns allows the investor to predict with some certainty a return that the fund will have. Although much of the variability will be unpredictable, by predicting a small amount of the variability of the mean return, it may be possible to substantially change the mean return of the fund, perhaps even so much as to double the mean return. In summary, the mean returns are indicative of future relative performance, if properly adjusted. Based on this important point in evaluating the results, perhaps prospectuses should include information on how to interpret past performance.

The answer to the question asked earlier about the validity of the disclaimer statement required in a prospectus is dependent on the definition of "indicative". If "indicative" is defined as past returns are an unbiased estimator of future returns, then the answer to the question would be no. Edwards and Ma (1988) showed that past returns are a biased estimator of future returns, thus showing that selectivity bias was present. However, if "indicative" is defined as past returns have some predictive ability in predicting future returns, than yes, the statement is mostly true. The measures of performance tested in this study do have some predictive ability. However, it should be pointed out that in relative terms, the results presented in this study seem to show that funds that perform better than other funds in the past tend to perform better than those same other funds in the future. Perhaps the disclaimer statement should be modified to reflect the specific nature of performance persistence found.

Overall, results indicate that a small amount of persistence is present in public, private, and CTA funds. Although the amount of persistence is small relative to the variation in the data, but large relative to the mean. The out-of-sample tests confirmed the results from the regression, a small amount of performance persistence. It is important to point out that this small amount of persistence makes it difficult to select the best single fund and therefore selecting a portfolio of funds is advisable. There is a possibility of picking the best set of funds, but precise methods and large amounts of historical data are required. In making these selections, the inclusion of a risk/return measure is necessary to give the investor as much information as possible to make informed investments in funds.

References

Beja, Avraham, and M. Barry Goldman. "On the Dynamic Behavior of Prices in Disequilibrium." *The Journal of Finance* 35(May 1980):235-248.

Brorsen, B. Wade, and Scott H. Irwin. "Futures Funds and Price Volatility." *Review of Futures Markets* 6(1987):118-135.

Brown, Stephen J., and William N. Goetzmann. "Performance Persistence." *The Journal of Finance* 50(June 1995):679-698.

Carhart, Mark M. "On Persistence in Mutual Fund Performance." *The Journal of Finance* 52(March 1997):57-82.

Clarke, Patricia. "Managed Futures: Distributional Characteristics, Return Seasonalities, and Implications for Performance Evaluation." unpub. manuscript. (November 1996).

Dowdy, Shirley, and Stanley Wearden. *Statistics for Research, Second Edition*. New York: John Wiley & Sons, 1991.

Edwards, Franklin R., and Cindy Ma. "Commodity Pool Performance: Is the Information Contained in Pool Prospectuses Useful?" *Journal of Futures Markets* 8(October 1988):589-616.

Elton, J., M. Gruber, and C. Blake. "The Persistence of Risk-Adjusted Mutual Fund Performance." *Journal of Business* 69(April 1996):133-157.

Elton, J., M. Gruber, and J. Rentzler. "Professionally Managed, Publicly Traded Commodity Funds." *Journal of Business* 60(1987):175-199.

Goetzmann, William, K., and Roger G. Ibbotson. "Do Winners Repeat?" *The Journal of Portfolio Management* 47(Winter 1994):9-18.

Goetzmann, William, K., and Brown. "Do Winners Repeat?" *The Journal of Portfolio Management* 47(Winter 1994):9-18.

Grossman, Sanford J., and Joseph E. Stiglitz. "On the Impossibility of Informationally Efficient Markets." *The American Economic Review* 70(June 1980):393-408.

Gruber, Martin J. "Another Puzzle: The Growth in Actively Managed Mutual Funds." *The Journal of Finance* 51(July 1996):783-810.

Hirt, Geoffrey A. and Stanley B. Block. *Fundamentals of Investment Management*, Fourth Edition. Boston: Richard D. Irwin, Inc., 1993.

Irwin, Scott H. "Further Evidence of the Usefulness of CTA Performance Information in Public Commodity Pool Prospectuses and a Proposal for Reform." *Advances in Futures and Options Research.* Greenwich: JAI Press Inc., 1994, pp. 251- 265.

Irwin, S. H., and B. W. Brorsen. "Public Futures Funds." *The Journal of Futures Markets* 5(Summer 1985):149-171.

Irwin, S., T. Krukemeyer, and C. Zulauf. "Are Public Commodity Pools a Good Investment?" *Managed Futures: Performance Evaluation and Analysis of Commodity Funds, Pools, and Accounts.* Chicago: Probus Publishing Company, 1992, pp. 403-434.

Irwin, S., C. Zulauf, and B. Ward. "The Predictability of Managed Futures Returns." *Journal of Derivatives* 2(Winter 1994):20-27.

Jensen, Michael C. "Risk, The Pricing of Capital Assets, and The Evaluation of Investment Portfolios." *The Journal of Business* 42(April 1969):167-247.

Kahn, Ronald N., and Andrew Rudd. "Does Historical Performance Predict Future Performance?" *Financial Analysts Journal* 51(November/December 1995):43-52.

Khorana, Ajay, and Edward Nelling. "The Performance, Risk, and Diversification of Sector Funds." *Financial Analysts Journal* 51(May/June 1997):62-73.

LaPorte Management ltd. Unpublished Data Sets on Managed Futures Funds. (August 1996).

McCafferty, Thomas A. *Winning with Managed Futures.* Chicago: Probus Publishing Company, 1994.

McCarthy, D., T. Schneeweis, and R. Spurgin. "Informational Content in Historical CTA Performance." *The Journal of Futures Markets* 17(May 1997):317-339.

Peters, Carl C. *Managed Futures.* Chicago: Probus Publishing Company, 1992.

Phelps, Shawn, and Larry Detzel. "The Nonpersistence of Mutual Fund Performance." *Quarterly Journal of Business Economics* 36(Spring 1997):55-69.

SAS Institute Inc., *SAS/IML© Software: Usage and Reference, Version 6,* First Edition, Cary, NC: SAS Institute Inc., 1989.

SAS Institute Inc., *SAS© Language: Reference, Version 6,* First Edition, Cary, NC: SAS Institute Inc., 1990.

SAS Institute Inc., *SAS© Procedures Guide, Version 6,* Third Edition, Cary, NC: SAS Institute Inc., 1990.

SAS Institute Inc., *SAS/STAT© User's Guide, Release 6.03,* Edition, Cary, NC: SAS Institute Inc., 1988.

Schwager, Jack D. *Managed Trading: The Myths and Truths.* New York: John Wiley & Sons, 1996.

Teweles, Richard J., and Frank J. Jones. *The Futures Game.* New York: Mcgraw-Hill Book Company, 1987.

Volkman, David A., and Mark E. Wohar. "Determinants of Persistence in Relative Performance of Mutual Funds." *The Journal of Financial Research* 18(Winter 1995):415-430.

Zweig, Jason. "Peering into a Fund's Past Can't Tell You All You Need to Know about Its Future." *Money* 25(October 1996):56.

Appendix A

Tables Listing Spearman Coefficients From Public, Private, and CTA Data Sets.

Table 1. Spearman's Coefficient between Selection α and Performance α and Mean α's: Combined CTA Data

Years in selection and performance period	r_s [a]	α_{sel} [b]	α_{per} [c]
1980-84	-0.063	1.913	-0.338
1981-85	0.345*[d]	1.467	1.120
1982-86	0.260	1.593	-0.106
1983-87	0.176	1.320	2.529
1984-88	0.056	1.789	1.662
1985-89	0.122	2.387	0.055
1986-90	0.229*	1.871	1.881
1987-91	0.053	1.992	0.760
1988-92	0.005	1.232	0.547
1989-93	0.071	0.997	1.252
1990-94	0.188*	1.037	0.192
1991-95	-0.080	0.682	1.086

[a]Spearman's Correlation coefficient.
[b]Mean of α's of selection period.
[c]Mean of α's of performance period.
[d]Asterisks indicate significance at .05 level.

Table 2. Spearman's Coefficient between Selection Returns and Performance Return and Mean Returns: Combined CTA Data

Years in selection and performance period	r_s	$mean_{sel}$[a]	$mean_{per}$[b]
1980-84	-0.070	1.913	-0.347
1981-85	0.335*[c]	1.465	1.191
1982-86	0.256	1.594	-0.095
1983-87	0.176	1.320	2.529
1984-88	0.059	1.790	1.663
1985-89	0.133	2.387	0.063
1986-90	0.230*	1.871	1.873
1987-91	0.052	1.994	0.738
1988-92	0.009	1.233	0.510
1989-93	0.083	0.997	1.242
1990-94	0.199*	1.035	0.183
1991-95	-0.042	0.677	1.112

[a]Mean returns of selection period.
[b]Mean returns of performance period.
[c]Asterisks indicate significance at .05 level.

Table 3. Spearman's Coefficients between Selection α/σ and Performance α/σ and Mean α/σ's: Combined CTA Data

Years in selection and performance period	r_s	$(\alpha/\sigma)_{sel}$[a]	$(\alpha/\sigma)_{per}$[b]
1980-84	0.060	1.462	-0.007
1981-85	0.330*[c]	1.209	0.658
1982-86	0.122	1.400	0.016
1983-87	0.217	1.091	1.293
1984-88	0.126	1.493	0.815
1985-89	0.165	1.776	0.052
1986-90	0.242*	1.419	1.314
1987-91	0.242*	1.726	0.589
1988-92	0.112	1.322	0.624
1989-93	0.168*	1.262	1.099
1990-94	0.235*	1.511	0.175
1991-95	0.002	1.084	1.092

[a]Adjusted Sharpe ratio for selection period.
[b]Adjusted Sharpe ratio for performance period.
[c]Asterisks indicate significance at .05 level.

Table 4. Spearman's Coefficient between Selection α and Performance α and Mean α's: Public Fund Data

Years in selection and performance period	r_s[a]	α_{sel}[b]	α_{per}[c]
1980-84	0.619*[d]	0.614	0.605
1981-85	-0.257	0.523	0.871
1982-86	0.007	0.686	-2.592
1983-87	0.054	0.307	2.236
1984-88	-0.057	0.924	0.624
1985-89	-0.442*	0.886	-1.150
1986-90	0.063	0.450	1.195
1987-91	0.291*	0.761	0.720
1988-92	0.274*	0.237	0.145
1989-93	0.166	0.275	0.673
1990-94	0.035	0.434	-0.547
1991-95	0.304*	0.090	0.959

[a]Spearman's Correlation coefficient.
[b]Mean of α's of selection period.
[c]Mean of α's of performance period.
[d]Asterisks indicate significance at .05 level.

Table 5. Spearman's Coefficient between Selection Returns and Performance Return and Mean Returns: Public Fund Data

Years in selection and performance period	r_s	$mean_{sel}$[a]	$mean_{per}$[b]
1980-84	0.575*[c]	0.605	0.604
1981-85	-0.279	0.515	0.849
1982-86	0.000	0.702	-2.604
1983-87	0.057	0.310	2.232
1984-88	-0.056	0.924	0.624
1985-89	-0.442*	0.886	-1.150
1986-90	0.063	0.450	1.195
1987-91	0.317*	0.762	0.730
1988-92	0.274*	0.237	0.145
1989-93	0.166	0.275	0.673
1990-94	0.030	0.435	-0.543
1991-95	0.306*	0.090	0.963

[a]Mean returns of selection period.
[b]Mean returns of performance period.
[c]Asterisks indicate significance at .05 level.

Table 6. Spearman's Coefficients between Selection α/σ and Performance α/σ and Mean α/σ's: Public Fund Data

Years in selection and performance period	r_s	$(\alpha/\sigma)_{sc}$[a]	$(\alpha/\sigma)_{per}$[b]
1980-84	0.658*[c]	0.729	0.395
1981-85	0.426	0.856	0.977
1982-86	-0.039	0.937	-0.895
1983-87	0.142	0.424	1.782
1984-88	0.076	1.240	0.819
1985-89	-0.246*	1.228	-0.523
1986-90	0.177	0.731	1.191
1987-91	0.261*	1.308	0.551
1988-92	0.320*	0.605	0.117
1989-93	0.205*	0.611	0.829
1990-94	0.015	0.952	-0.442
1991-95	0.434*	0.423	1.084

[a]Adjusted Sharpe ratio for selection period.
[b]Adjusted Sharpe ratio for performance period.
[c]Asterisks indicate significance at .05 level.

Table 7. Spearman's Coefficient between Selection α and Performance α and Mean α's: Private Fund Data

Years in selection and performance period	r_s[a]	α_{sel}[b]	α_{per}[c]
1980-84	-0.127	1.824	1.360
1981-85	-0.182	1.224	1.384
1982-86	-0.027	1.562	0.007
1983-87	0.168	1.364	2.549
1984-88	-0.097	1.492	1.323
1985-89	0.298	1.651	0.022
1986-90	0.253	1.677	1.676
1987-91	0.019	1.564	0.697
1988-92	0.115	0.794	0.388
1989-93	0.120	0.774	0.998
1990-94	-0.030	0.937	0.203
1991-95	0.056	0.420	1.061

[a]Spearman's Correlation coefficient.
[b]Mean of α's of selection period.
[c]Mean of α's of performance period.

Table 8. Spearman's Coefficient between Selection Returns and Performance Return and Mean Returns: Private Fund Data

Years in selection and performance period	r_s	$mean_{sel}$ [a]	$mean_{per}$ [b]
1980-84	-0.079	1.961	1.360
1981-85	-0.182	1.282	1.442
1982-86	-0.027	1.557	-0.148
1983-87	0.202	1.361	2.407
1984-88	-0.033	1.492	0.979
1985-89	0.324* [c]	1.666	-0.051
1986-90	0.288*	1.676	1.720
1987-91	0.013	1.565	0.718
1988-92	0.114	0.794	0.400
1989-93	0.149	0.774	1.029
1990-94	-0.008	0.937	0.194
1991-95	0.056	0.419	1.067

[a]Mean returns of selection period.
[b]Mean returns of performance period.
[c]Asterisks indicate significance at .05 level.

Table 9. Spearman's Coefficients between Selection α/σ and Performance α/σ and Mean α/σ's: Private Fund Data

Years in selection and performance period	r_s	$(\alpha/\sigma)_{sel}$ [a]	$(\alpha/\sigma)_{per}$ [b]
1980-84	0.406	3.021	1.466
1981-85	0.464	2.272	2.349
1982-86	-0.038	2.612	0.601
1983-87	0.430*[c]	1.575	2.186
1984-88	0.319	1.723	1.448
1985-89	0.345*	1.523	0.468
1986-90	0.586*	1.538	1.275
1987-91	0.228	1.867	0.691
1988-92	0.199	1.251	0.907
1989-93	0.298*	1.244	1.300
1990-94	0.315*	1.767	0.657
1991-95	0.310*	1.164	1.320

[a]Adjusted Sharpe ratio for selection period.
[b]Adjusted Sharpe ratio for performance period.
[c]Asterisks indicate significance at .05 level.

VITA

John P. Townsend

Candidate for the Degree of

Doctor of Philosophy

Thesis: PERFORMANCE PERSISTENCE IN MANAGED FUTURES FUNDS

Major Field: Agricultural Economics

Biographical:

Personal Data: Born in Safford, Arizona, on May 5, 1969, the son of Jerry and Margo Townsend.

Education: Graduated from Moriarty High School, Moriarty, New Mexico in May 1987; received Bachelor of Science degree in Agricultural Extension and Education and a Master of Science degree in Agricultural Economics from New Mexico State University, Las Cruces, New Mexico in May 1992 and May 1994, respectively. Completed the requirements for the Doctor of Philosophy degree with a major in Agricultural Economics at Oklahoma State University in July, 1998.

Professional Memberships: American Agricultural Economics Association.