APPLICATION OF BOOTSTRAP METHODOLOGY IN FUND PERFORMANCE

ELİF ÇAKMAKOĞLU
106626008

İSTANBUL BİLGİ UNIVERSITY
INSTITUTE OF SOCIAL SCIENCES
FINANCIAL ECONOMICS – QUANTITATIVE FINANCE

PROF. DR. BURAK SALTOĞLU
2010

Application of Bootstrap Methodology in Fund Performance

Elif Çakmakoğlu
106626008

Prof. Dr. Burak Saltoğlu:
Assoc. Prof. Dr. Ege Yazgan
Asst. Prof. Dr. Orhan Erdem

Date of Approval: June 9, 2010

Number of Pages: 98

Keywords (English)	Keywords (Turkish)
1) Bootstrap	1) Bootstrap
2) Alpha	2) Alfa
3) Performance Measurement	3) Performance Ölçümü
4) Luck	4) Şans
5) Skill	5) Yetenek

Abstract

This study investigates any evidence of superior performance among 65 Turkish funds over the period February 2004 to December 2009, as well as any evidence of luck. Bootstrap procedure is used to construct a distribution of the t statistics of performance measure alpha with the null hypothesis of zero abnormal performance. With this procedure, any abnormal performance resulting from luck is separated from the actual. The results indicate that abnormal performances of Turkish funds are due to skill rather than luck. Furthermore, bootstrap simulation of Henriksson and Merton timing model reveals that the inability of timing within Turkish fund managers diminishes good performance resulting from the stock picking ability of them.

Özet

Bu çalışma Türk piyasasından seçilen 65 fonun Şubat 2004 ve Aralık 2009 tarihleri arasında yüksek performans gösterip göstermediğini ve şans faktörünün varlığını araştırmaktadır. Sıfır olağandışı performans boş hipotezi ile alfanın t istatistiklerinin dağılımının oluşturulmasında bootstrap yöntemi kullanılmıştır. Bu yöntem ile şansa dayalı olağandışı performans gerçekleşenden ayrıştırılmaktadır. Sonuçlar, Türk fonlarının olağandışı performansının şanstan ziyade yetenekten kaynaklandığını işaret etmektedir. Ayrıca, Henriksson ve Merton zamanlama modelinin bootstrap simülasyonu Türk fon yöneticilerinin zamanlama hatalarının, hisse seçme becerilerinden kaynaklanan iyi performansı yok ettiğini ortaya çıkartmaktadır.

Acknowledgment

I would like to show my gratitude to my supervisor Prof. Dr. Burak Saltoğlu who provided me the necessary sources and also put me in the right track whenever I needed. Without his guidance I would never be able to complete this study. I am also indebted to Serhat Yücel who studied with me in all stages of my thesis. His support and encouragement motivated me to finish this work. I am grateful to Demet Demir for her patience and kindness in finalizing the administrative procedure. She has made things a lot easier for me. Finally I am heartily thankful to my family and friends for their moral support. They bared me through tough times.

Table of Contents

List of Tables

List of Figures

Abbreviations

ISE	Istanbul Stock Exchange
EFAMA	European Fund and Asset Management Association
USA	United States of America
CAPM	Capital Assets Pricing Model
SMB	Small Minus Big Sized Firms
HML	High Minus Low Book-to-Market Equity Ratio Firms
PR1YR	Difference of One Month Lagged 11 Month Return of Best and Worst %30 Firms' Average Return
CMB	Capital Markets Board of Turkey
UK	United Kingdom
ISE	Istanbul Stock Exchange
SML	Security Market Line
ST1	Strateji Menkul Değerler A.Ş. A Tipi Değişken Fon
MAD	Yatirim Menkul Değerler A.Ş. A Tipi Değişken Fon
NYSE	New York Stock Exchange
BE/ME	Book-to-Market Equity Ratio
KYD30	30 Days Bond Index of Turkish Institutional Investment Managers
OLS	Ordinary Least Squares

1. Introduction

This study analyses the performance of 65 Turkish funds for the period February 2004 to December 2009 and tries to separate luck from skill using bootstrap procedure. Throughout the world there are various studies trying to measure the performance of mutual and pension funds. In these studies performance is decomposed into two components. One is called the timing component which is the ability of the fund manager to forecast the general behavior of the market prices and the other component is called the stock picking ability which is the ability of the manager to invest in exceptionally performing individual funds.

Jensen's alpha is the most famous stock picking measure. Since the sixties, various regression models are developed for the appropriate alpha to measure the stock picking ability. Choosing the right model and discussing whether the fund managers have any stock picking ability is a popular topic held in detail. However, whether this alpha is due to true stock picking ability or chance is rather a new topic. It is first addressed by Kosowski, Timmermann, Wermers and White (2005) with the bootstrap methodology for US mutual funds and later by Cuthbertson, Nitzsche and O'Sullivan (2004) for UK equity mutual funds. According to our knowledge, this is the first study trying to separate luck from skill in Turkey.

By implementing bootstrap methodology for fund performance we try to improve the validity of the alpha measures. Historically well known performance measures have the assumption of normality for performance measure alpha meanwhile for a high percentage of funds in US and UK it is shown to be non-normal by Kosowski et al and Cuthbertson et al, respectively. Varying risk levels of the funds also violates the assumption of normality for the joint distribution of the funds. For these purposes, we apply bootstrap procedure to the selected Turkish funds and try to create an environment where we can have an idea about the level of alphas that can be achieved solely by chance. Then, we compare actual alphas with our bootstrap results and try to decide the level of stock picking ability for Turkish fund managers.

Funds that are included in this study are the ones that existed throughout the observation period. Excluding funds that do not satisfy this criterion may cause survivorship bias. To make sure the results that we obtain are valid, several survivorship bias measures are computed in light of the survivorship bias measure of Blake and Timmermann (1998).

This paper proceeds as follows: Section 2 gives a brief summary of the fund management environment and history for Turkish markets and the comparison of it with the international ones. Then, the first four parts of section 3 discusses the literature of fund performance and presents the well known models of performance and the last two parts presents the resampling methodologies and several bootstrapping techniques. Section 4 presents our data and discusses the issue of survivorship bias which is a bias resulting from the exclusion of funds from the dataset. Further detail on the well known performance measures are given in section 5.1 together with the estimation results. Then resampling methodologies and bootstrap procedure is introduced and the reasoning behind using it is discussed in section 5.2. Empirical findings of the bootstrap procedure for several models are demonstrated in section 6. Finally section 7 concludes the study.

2. Developments in Mutual and Pension Fund Industry in Turkey

Mutual fund is an investment product operated by professional managers that uses the money collected from many investors in return for participation certificates to invest in investment securities such as stocks, bonds, real estate, money market instruments, commodities, precious metals, other mutual funds and securities, with fiduciary ownership principals and diversification of risk. Mutual funds are not legal entities, but to protect the rights of the investor the assets held within a mutual fund are separately held from the founder of the fund. Founder of the mutual fund is responsible for the protection and the safekeeping of the assets of the fund. Mutual funds are required to protect the assets under management by depositing them in a depository institution (the ISE Settlement and Custody Bank, Inc). This way it is ensured that the mutual fund assets aren't affected from the state of the founder, such as bankruptcy[1]. Moreover assets of a fund cannot be pledged, used as collateral or seized by a third party, which fortifies the rights of the investor. Mutual funds in Turkey can be founded by banks, insurance companies, pension companies, intermediary institutions and eligible retirement and provident funds[2]. They are managed by either intermediary institutions or portfolio management companies with certain legal restrictions ensuring that the internal statute of the fund is met. There are certain limitations in the regulation of mutual funds in order to assure sufficient liquidity and proper diversification against risks taken. Mutual funds can invest at most 10% of their assets under management into the securities issued by a single issuer. Mutual funds can invest in at most 9% of any issuer's total shares. Funds managed by the same manager can invest in at most 20% of any issuer's shares in total. Mutual funds can invest at most 10% of their total shares into shares or dept instruments that are not quoted to a stock exchange. Mutual funds can invest at most 5% of their assets under management into securities that are issued by the founder or the fund manager with the limitation that the amount invested in should at most be 10% of the issued security. Mutual funds cannot be represented in the management of the companies whose shares it has purchased. Mutual funds cannot short sell[3]. Mutual funds should maintain the fair price - buying from the lowest possible price and selling from the highest possible price, in the transactions of the fund.

Turkish mutual funds are classified into two types which are called "Type A" and "Type B". Type A funds are required to invest at least 25% of their total holdings on monthly average basis into Turkish securities issued by Turkish companies. All other mutual funds are classified as Type B.

Mutual funds are classified into different categories according to their investment objectives. Funds that invest at least 51% of their total assets for all times in public and private debt instruments are called “bonds and bills funds”. “Stock funds” are funds that invest at least 51% of their assets under management permanently in securities issued by Turkish companies. “Sector funds” are the ones that invest at least 51% of their assets permanently in securities of companies belonging to the same sector. Funds that invest 51% of their assets under management for all times in securities issued by subsidiaries are called “subsidiary funds”. “Group funds” are funds that invest at least 51% of their assets permanently in securities issued by a certain group. “Foreign securities funds” invest at least 51% of their assets for all times in foreign public or corporate sector securities. Funds that invest 51% of their assets under management permanently in gold and other precious metals or capital market instruments backed by these metals traded in national and international stock markets are called “precious metals funds”. Funds that invest at least 51% of their assets under management for all times in gold and capital market instruments backed by gold in national and international exchanges are called “gold funds”. “Composite funds” are funds that invest all their assets in at least two of the instruments like shares, dept instruments, gold and other precious metals or assets backed by them with the restriction that the share of each instrument is not less than 20% of the total fund assets. “Liquid funds” are funds that invest all their assets continuously in liquid capital market instruments with less than 180 days to maturity with the requirement that the weighted average maturity of the portfolio being at most 45 days. Funds that cannot be classified in one of the above mentioned categories with respect to portfolio limitations are called “variable funds”. “Index funds” are funds that invest at least 80% of their assets continuously in an index or shares held within an index with correlation coefficient of at least 90% between the unit share value of the fund and the value of the index. “Funds of funds” invest at least 80% of their assets continuously in other funds. “Guaranteed funds” are umbrella funds aiming a return over the total or part of the startup investment within a specified horizon with a guarantee of the guarantor and with an appropriate investment strategy. “Protected funds” are umbrella funds with the intension of a return over the total or part of the startup investment within a specified horizon with the best effort investment strategy. Finally funds with participation certificates given to certain individuals or institutions are called “special funds’[4].

Liberalization in Turkish economy began in early 1980’s. First move was the acceptance of the Capital Markets Law in 1981. Four years after the Capital Markets Law, in December 1985, Istanbul Stock Exchange was founded and began trading in 1986 in Cağaloğlu office. One year later first communiqué regulating the mutual funds was formed and in July 1987, the first mutual fund was established by İşbank. This B type liquid fund is still alive. Also in 1987, Istanbul Stock Exchange moved to its modern headquarters, which is the Karaköy office. In July 1992, new regulations consisting of the types of mutual funds were formed and A type funds were exempted from tax in order to encourage investments in stocks. Later on, extensive and innovative communiqué was launched in 1996 which is still in use. First private mutual fund, foreign mutual fund and sector fund were issued in 1997 followed by the issuance of the first index fund in 1999. In 2001 the first communiqué regulating the pension funds was published and in 2003 six pension fund companies started operating. In 2007 insurance law was published in the official gazette and with this law, several changes and extensions occurred in the communiqué regulating the pension funds. In order to secure the rights of the investors Pension Monitoring Center controls the operation of the pension fund system. In 2007 the responsibilities and authorities of the Pension Monitoring Center were redefined and several changes in the communiqué were made. Last revisions regarding the communiqué were made in 2008.

Some argue that the ancestors of modern funds were formed by a Dutch merchant in late 18th century. Others argue that it was formed in 1882 by the King of Netherlands. Although the roots are not certain first modern mutual fund was the Massachusetts Investors' Trust which was established in 1924 and went public in 1928. Number of mutual funds is increasing since 1930's especially escalating in 1980's and 1990's. The total mutual fund industry by the end of 2009 is 15.9 trillion euros. Out of this amount 55% is managed in Americas and 33% is managed in Europe. Largest single markets are USA, Luxemburg and France with investment assets amounting to 7.7 trillion euros, 1.6 trillion euros and 1.25 trillion euros, respectively. Turkey with investment assets of 13.4 billion euros has a market share of 0.085%. The most popular type of fund in both USA and Europe is the equity fund. Invested assets for equity funds are 3.4 trillion euros and 1.58 trillion euros, respectively. According to EFAMA statistics worldwide investment fund assets decompose as follows; 39% are invested in equity funds, 23% in money market funds, 20% in bond funds, 10% in balanced-mixed funds and 8% in other funds. By the end of 2009, there are 65306 mutual funds operating worldwide. Out of these, 33054 is located in Europe, 286 of them being in Turkey, 16553 in Americas and 14795 in Asia and Pasific countries.

Services provided by mutual funds cannot be limited to any timing or stock picking ability. Investing in mutual funds has certain advantages for small investors. First of all, mutual funds, manage huge amounts gathered by small investors. Thus, a mutual fund can easily invest in certain capital markets, derivative markets or international markets where small investors do not have access to. Further, due to the size of the total amount managed, or the bargaining skills of the professional managers hired, mutual funds can easily bargain and get better prices compared to small investors. Secondly, a mutual fund can keep more instruments in its portfolio compared to a small investor's and hence provides diversification to the customers. In most countries including Turkey, mutual funds are constructed as separate entities where any distress like default or bankruptcy of the parent company is isolated from the fund. Furthermore, the assets of the fund are kept in custody to ensure that mutual funds are legally safe investment alternatives. In addition to these advantages, mutual funds are supported by the government where they are permanently exempted from the corporate tax. A mutual fund can also provide different strategy alternatives to small investors where they cannot replicate, given the time and resource limitations. Index funds are example to strategies provided.

With all these advantages, mutual funds deduct management fees for the services provided, which is a negative effect compared to a similar portfolio constructed by the investor himself. Although the fund manager is not planning to change his position, due to incoming and outgoing cash flows, he might need to do trading, which creates transaction costs.

3. Review of Literature

Bootstrap is gaining more and more popularity in recent years due to heterogeneous risk taking of fund managers and the non-normal distribution of individual fund alphas. Furthermore, bootstrap provides a powerful tool for performance persistence testing compared to parametric t-tests over past performance data. One of the most popular performance measures used in bootstrap procedures is the alpha.

3.1 Jensen's Alpha and Other Popular Measures of 1960's.

Alpha is the selectivity measure of Jensen (1967) which is the intercept of a one factor regression model, excess return of the fund regressed to the excess return of the market.

$$\tilde{R}_{jt} - R_{ft} = \alpha_j + \beta_j\left[\tilde{R}_{Mt} - R_{ft}\right] + \tilde{u}_{jt} \qquad (1)$$

For time t, $\tilde{R}_{jt}$ is the return of portfolio j, $\tilde{R}_{Mt}$ is the return of the market portfolio and R_{ft} is the risk free rate. If the manager has any selection ability independent from the state of the market, the regression line should shift upwards in parallel with any other ordinary portfolio manager. This upward shift causes the intercept (which was originally zero in the CAPM) to have a positive value which is represented by α_j. One of the most important aspects of Jensen's alpha is that it is an absolute measure rather than a ranking measure. Jensen has shown the bias of the least squares estimator $\hat{\beta}$. According to him, beta is downward biased proportional to the manager's forecasting ability of the market. Hence if the manager doesn't have any timing ability of the market, then $\hat{\beta}$ will be unbiased and so will $\hat{\alpha}$. But if the manager is capable of forecasting the market then $\hat{\beta}$ will be downward and $\hat{\alpha}$ will be upward biased. But later Grant (1977) has shown that the bias caused by $\hat{\beta}$ differs from Jensen's both by magnitude and direction. Besides other advantages of investing in mutual funds (for example; diversification) Jensen investigated the stock picking ability. He examined 115 mutual funds for the period 1945 to 1964. He found very little evidence of stock picking ability. The results hold even for gross returns.

Treynor (1965) had developed a return over risk ranking measure. The most important aspect of the measure is how it defines the risk. Treynor assumes that risk component should not get affected from market fluctuations and should only reflect the variability of the fund compared to the market. So, the definition of risk for Treynor is the beta coefficient of the linear one factor regression of excess fund return over the excess market return.

$$T = \frac{R_j - R_f}{\beta_j} \qquad (2)$$

Treynor measure gives an idea about the extra return gained against one unit of risk. One drawback compared to Jensen's measure is that Treynor's measure is a ranking measure rather than an absolute measure.

Sharpe compared to Treynor used the volatility of a fund as the measure of risk and developed the "reward to variability ratio (R/V)".

$$R/V = \frac{R_j - R_f}{\sigma_j} \qquad (3)$$

If the aim is to evaluate the past performance of a fund, the R/V ratio will be a better measure compared to Treynor ratio since, it includes the total variability. On the other hand, for predicting the future performance, Treynor ratio will be a better estimator since it is net of market variability and lets us identify the deviation from the market.

3.2 Performance Measures of Stock Picking Ability

Alpha is a measure of stock picking ability. But there are certain generally accepted trading strategies depending on firm characteristics which can beat the market but still, has nothing to do with any stock picking ability. For example one can beat the market with simply certain passive investment strategies like investing into high book-to-market equity ratio stocks. Fama and French (1993 and 1996) have investigated similar patterns and added them as independent variables to the factor model in order to purify alpha from this noise.

$$\tilde{R}_{jt} - R_{ft} = \alpha_j + \beta_{1j}\left[\tilde{R}_{Mt} - R_{ft}\right] + \beta_{2j}(SMB_t) + \beta_{3j}(HML_t) + \tilde{e}_{jt} \qquad (4)$$

SMB_t (small minus big) is the index generated by Fama and French in order to capture the return of passive investment strategies related to size and HML_t (high minus low) is the index used for strategies related to book-to-market equity ratio.

Jegadeesh and Titman (1993) documented that portfolios buying stocks with high past returns and selling stocks that have low past returns generate significant positive returns in the short run. Moreover, they have expressed that the significant returns deteriorates in three years time. Furthermore, Daniel, Grinblatt, Titman and Wermers' (1997) findings suggest that there is a certain momentum effect in stock returns.

Affected by the work of Jegadeesh and Titman (1993) and Daniel et al (1997), Carhart (1997) developed Fama and French (1993 and 1996) model with a momentum factor.

$$\tilde{R}_{jt} - R_{ft} = \alpha_j + \beta_{1j}\left[\tilde{R}_{Mt} - R_{ft}\right] + \beta_{2j}(SMB_t) + \beta_{3j}(HML_t) + \beta_{4j}(PR1YR_t) + \tilde{e}_{jt} \quad (5)$$

Along with size and book to market equity, $PR1YR_t$, is the added factor which is the one year momentum in stock returns.

Carhart, found that funds with good performance last year tend to have above average returns next year, but this above average return decays later on.

3.3 Performance Measures of Market Timing

Performance measures discussed previously were in search of stock picking ability. For the sake of a better measurement, they have tried to identify passive strategies which might affect alpha and added them as factors to the regression model. Besides stock picking ability, another important talent that the theoreticians tried to identify is the market timing ability. Market timing ability is the ability of the fund manager to identify the trends and update his position accordingly. Beta of a linear regression is the factor that measures the leverage of the market. So, if the manager has any timing ability he will decrease his beta in bad times and increase his beta in good times in order to optimize the effect of fluctuations of the market on his portfolio.

One of the well known efforts of market timing ability is the Treynor and Mazuy (1966) model. Treynor and Mazuy searched for any market outguessing ability for 57 mutual funds during the period beginning in 1953 and ending in 1962. They defined the characteristic line as the fitted line between the market returns and the fund returns. Capital asset line is the sensitivity of the fund returns to the market returns. Assuming that the fund manager has any timing ability he will shift to less volatile securities (which means lower beta) if he anticipates that the market is going to fall and he will shift to more volatile securities (which means higher beta) if he thinks that the market is going to rise. Since no manager claims to be able to predict the market perfectly Treynor and Mazuy assumes that the beta will increase as the market increases and decrease as the market decreases. High beta for good times and low beta for bad times condition in practice occurs as a beta smoothly increasing or decreasing according to market conditions rather than two lines with an elbow, which means a concave upwards capital asset line. To capture this convexity, Treynor and Mazuy added the square of the excess market return to the regression.

$$\tilde{R}_{jt} - R_{ft} = \alpha_j + \beta_j[\tilde{R}_{Mt} - R_{ft}] + \gamma_j[\tilde{R}_{Mt} - R_{ft}]^2 + \tilde{e}_{jt} \qquad (6)$$

γ_j captures the timing ability of the manager. They empirically found that there is only one fund with market timing ability which can be explained by luck so; there is no evidence of market timing ability. Hence, portfolio managers shouldn't try to time the market.

Henriksson and Merton (1981) assumes that the manager can only forecast whether the market return exceeds risk free rate or not and doesn't have any information of the magnitude of it. So instead of squared market excess return as in Treynor and Mazuy (1966), they used a dummy variable *D* indicating the two states of the world.

$$\tilde{R}_{jt} - R_{ft} = \alpha_j + \beta_j[\tilde{R}_{Mt} - R_{ft}] + \gamma_j\left[[\tilde{R}_{Mt} - R_{ft}]D\right] + \tilde{e}_{jt} \qquad (7)$$

$$D = \begin{cases} 0 & \tilde{R}_{Mt} \leq R_{ft} \\ 1 & \tilde{R}_{Mt} > R_{ft} \end{cases} \qquad (8)$$

A statistically significant and positive γ_j will mean that the portfolio manager has market timing ability.

3.4 Conditional Performance Measures

Until now we have discussed unconditional models of performance. But in reality, alpha and beta values may be conditional on a specific information set and they may be time varying. Thus, in a world where the alpha or the beta is conditional on any information set, measuring performance over unconditional averages will give us to misleading results.

Ferson and Schadt (1996) have conditioned the beta to certain market information. They have assumed that beta, $\beta_{jm}(Z_t)$ consists of two components: the average long term beta, β_{0j} and beta conditioned on market information, $\beta_{uj}{}'$.

$$\beta_{jm}(Z_t) = \beta_{0j} + \beta_{uj}{}' z_t \qquad (9)$$

Here, z_t is the vector of the deviations from the expectations of the information Z_t that the beta is conditioned on. The conditional beta definition of Ferson and Schadt can be implemented into any regression model. For example the conditional form of eq. (1) will be as follows:

$$r_{jt+1} = \alpha_j + \beta_{0j} r_{mt+1} + \beta_{uj}^{'}(z_t r_{mt+1}) + e_{jt+1} \qquad (10)$$

where $r_{jt+1} = \tilde{R}_{jt+1} - R_{ft+1}$ and $r_{mt+1} = \tilde{R}_{Mt+1} - R_{ft+1}$. Ferson and Schadt showed empirically that conditioning the performance measures is statistically significant. They have also concluded that the inferior alpha performance is due to using unconditional average information and the distribution of alpha shifts through zero after conditioning the beta to market information. Furthermore, they have also expressed conditional versions of Treynor and Mazuy (1966) and Henriksson and Merton (1981) better predict market timing ability.

Christopherson, Ferson and Glassman (1998) have further developed the conditional model of Ferson and Schadt with conditioning the alpha as well as beta.

$$r_{jt+1} = \alpha_{0j} + A_j^{'} z_t + \beta_{0j} r_{mt+1} + \beta_{uj}^{'}(z_t r_{mt+1}) + e_{jt+1} \qquad (11)$$

Where α_{0j} is the average alpha and $A_j^{'}$ is the alpha that is time varying, conditioned on the information set Z_t. They have analyzed pension fund data and found that bad performance persists.

3.5 Bootstrap and Other Resampling Approaches

Bootstrap is a simulation based statistical analysis which is an alternative to the traditional statistical techniques that assume certain probability distribution. It is a process of resampling the data at hand to build a sampling distribution. The research on this topic began in the late 1970s although early work that influenced bootstrapping traces back to 1920s to R. A. Fisher's work on maximum likelihood estimation (Efron, 1998). Fisher's method is not only practical but also looses less information in small samples. Smart mathematics he uses was an advantage before the technological developments, now bootstrap has the advantage in this sense. The similarity of Fisher information with bootstrap is the substitution of the estimates for the unknown parameter. There are also other methods, resampling methods that are related to the history of bootstrapping such as jackknifing, cross-validation, random subsampling and permutation tests.

Jackknife was first introduced by Quenouille (1949). Jackknife statistics are produced by dropping out data from the sample one at a time and calculating the necessary statistics over that new sample. This technique lost its popularity to bootstrapping, since it is more generalized, but jackknife is easier to apply to complex sampling.

Cross-validation is another resampling method that predates bootstrapping. "Leave one out" cross-validation is usually confused with jackknifing. Both techniques omit one observation at a time and work on the remaining subset. But while jackknifing is used to estimate the bias of a statistic, cross-validation is used to estimate the prediction error. Cross-validation is a way of measuring the prediction accuracy of different models and selecting the one with the smallest prediction error. K-fold cross-validation is done simply by dividing the observation set into k equal (or near equal) segments, leaving one segment for testing and working with the other k-1 segments for building the model. Bootstrapping has small variance in small samples while cross-validation nearly provides unbiased estimates of prediction error. A type of bootstrapping (.632+ bootstrap) outperforms cross-validation according to Efron and Tibshirani (1995). Random subsampling validation is done by choosing repeated random sets for validation and training. This repeated randomness causes some observations to appear more than once and some not at all in the validation set which biases the results.

Another resampling method used in the literature is the permutation test which is based on the work of R.A. Fisher in the 1930s. It checks the significance of a statistic by forming a distribution of it from all possible rearrangements of the sample. It is commonly used for testing whether the distributions of two samples are equal, while bootstrapping is used for parameter estimation. Because of this specific hypothesis, permutation test has a limited use. Compared with the parametric methods, both permutation test and bootstrapping have the advantage of finding results without making parametric assumptions. But when applicable, permutation test gives exact results while bootstrapping gives only an approximation.

Most of the developments on the bootstrap theory occurred after Efron (1979). Eventhough this new technique didn’t grab researchers’ attention at first, Efron’s continuous publications on the topic created the deserved curiosity. Bickley and Freedman (1981) and Singh (1981) tested the accuracy of the technique on several ocations and showed that the bootstrapping works for the sample mean when there are finite second order moments. Bickley and Freedman (1981) also gave some examples for situations where bootstrapping failed. There are many other research papers about the consistency of bootstrap technique (such as Athreya (1987) and Gine and Zinn (1989)) but there is still a lot to search, especially for the complex samples. There were alot of suspicions regarding the bootstrap methodology at first. The work of Efron and Tibshirani (1986) was a somewhat successful attempt to clarify the minds of scientists, but as Chernick (1999) explains in his book, oversimplicity of the expressions and the lack of mathematical theory resulted misunderstandings about the methodology.

After asymptotic consistency of bootstrapping became a popular research topic, the limitations and possible applications of it became clearer. Researches realized that bootstrapping isn't only used for estimation of the standard errors, confidence intervals and hypothesis tests but it can be used for a wide class of problems such as regression problems, time series analysis and density estimation as well as a wide class of fields such as geology, medicine, engineering, biology, psychology and econometrics. Having no assumptions about the distribution of the sample not only made bootstrapping applicable to many situations but also made things a lot easier. The simplicity of the technique made it more attractive among scientists, especially with the growing developments on computer technology. The idea with basic bootstrap is always to construct a sample with replacement from the original sample, compute the necessary statistics for the bootstrapped sample, repeat these steps many many times and obtain a distribution of the bootstrap statistics. You can then compare your results with the original statistics. The important thing is to be careful about the limitations of bootstrap and avoid using it when there is evidence of theoretical drawback. It is believed that bootstrapping gives acceptable results when all other approaches are eliminated with the assumptions of the sample.

There are various types of bootstrapping some of which are basic non-parametric bootstrap, parametric bootstrap, smooth bootstrap and moving block bootstrap. The basic bootstrap technique makes no assumptions about the population distribution and assumes that the sample at hand is a good representative of the population and drives conclutions using the sample. Therefore it is actually a non-parametric technique. If there is an assumed parametric distribution for the population, the estimate of interest can be calculated using parametric bootstrap method. In parametric bootstrap, samples are drawn from the parametric estimate of the population rather than using resampling with replacement from the sample. Other than this, the procedure proceeds similar to basic bootstrap. For each bootstrap iteration the relavent statistic is calculated and using these values bootstrap distribution is obtained. According to Efron and Tibshirani (1993), parametric bootstrap gives more accurate answers than analytic formulas and in the state of lacking these formulas it can provide answers. It is useful when some information about the population is available. It helps to deduce conclusions without any use of complicated formulas. It is not very common to use parametric bootstrap since the distribution of the population is known but some statistics are not easy to calculate and parametric bootstrap saves us from the trouble.

Smoothed bootstrap like parametric bootstrap doesn't use sampling with replacement from the sample. It rather uses sampling from a smooth estimate of the population. Efron (1982) applied smoothed bootstrap to estimate the standard error of the correlation coefficient using both Gaussian and uniform kernel functions and the results indicated that smoothed bootstrap estimations were somewhat better than the basic non-parametric bootstrap results. Silverman and Young (1987) found some conditions to use smoothed bootstrap procedure.

Basic non-parametric bootstrap method assumes that the observations are independent but this may not be the case for time series data or other correlated data. Moving block bootstrap is applied to correlated data. The sample is divided into b nonoverlapping blocks of l consecutive observations such that bl gives approximately the sample size. Bootstrap sample is constructed by randomly sampling b blocks with replacement and linking them together. This method is used to preserve the correlation within the data. The correlation between the observations is assumed to be strongest within the blocks and weaker between them. So the size of the blocks, l is important. If l is large, then the number of blocks (b) is small so the bootstrap samples will mostly be the same. If l is small, then observations in different blocks may not be independent which will reduce the accuracy of the inferences (blocks are assumed to be independent while bootstrapping). l should be choosen so that the observations that are l units apart from each other are nearly independent. This way the correlation present within a block is preserved. Moving block bootstrap is introduced by Carlstein (1986) and Künsch (1989) and then discussed by Efron and Tibshirani (1993). If the block length is choosen correctly it provides a simple alternative to parametric time series models preserving the empirical distribution and the correlation of the original sample.

Another application field of bootstrapping is the regression models. Regression models are useful tools in sorting out the effects of certain explanatory variables on a response variable. Ordinary least squares (OLS) estimation technique is commonly used for estimating the coefficients of the explanatory variables. But it is a good proxy only if the underlying assumptions are satisfied. Especially in small samples, the extreme values may cause the normality assumption of the residuals to be violated. The residuals may also have heavy-tailed distributions rather than normal. Both these factors cause least squares estimation to be misleading. Bootstrap procedure helps to manage these problems. Efron (1982) introduced two possible procedures for regression models, one of which is bootstrapping the pairs. With this procedure the explanatory variables of the regression model are treated as random. Vectors of the response variables with the corresponding explanatory variables are constructed and the resampling methodology is applied to these vectors with replacement. These vectors are then used to fit the model and obtain bootstrapped regression coefficients. The estimated bootstrap distribution of each regression coefficient is formed using equal probability.

The other bootstrap procedure that Efron (1982) proposed is called bootstrapping the residuals. This method treats the explanatory variables as fixed and uses the residuals instead. It is more complex compared to basic non-parametric bootstrapping since an underlying model should be structured in order to obtain the residuals. So the first step is to fit a model to the sample and obtain the observed residuals by simply subtracting the multiplication of the explanatory variables with the estimated coefficients from the response variable. After this procedure, for each bootstrap replication resample the residuals with replacement and adding these bootstrapped residuals to the multiplication of the explanatory variables with the estimated coefficients, obtain the bootstrapped response variables. Then regressing bootstrapped response variables against the explanatory variables, estimate the bootstrapped regression coefficients. As always these coefficients are used to form bootstrap distributions of the regression coefficients.

Since bootstrapping the residuals preserve the information coming from the explanatory variables, the choice of the model gains importance. If the underlying regression model isn't the best choice, bootstrapping the residuals may give misleading results. Efron claims that bootstrapping the pairs is less sensitive to this kind of faulty model selections. Therefore if there are doubts about the choice of the underlying model, it is better to use bootstrapping the pairs approach. Efron and Tibshirani (1986) assert that the results obtained from bootstrapping the pairs approach approximates to the results obtained from bootstrapping the residuals approach. But these methods differ for small samples therefore one should be careful dealing with them. Bootstrapping doesn't have strict boundaries. In some instances, even though the explanatory variables are random, it may be best to use bootstrapping the residuals approach and act as if they are fixed. The underlying model doesn't have to be a perfect fit for bootstrapping the residuals to give acceptable results.

The next section deals with two applications of bootstrapping the residuals approach used for fund performance.

3.6 Applications of Bootstrap Method to Fund Performance

Kosowski, Timmermann, Wermers and White (2005) tried to distinguish luck and skill in the performance measure of alpha. The pioneering work of Kosowski et al is the first paper that uses bootstrap technique for this purpose. Furthermore, they try to reconstruct data by bootstrapping to test for persistency rather than analyzing past performance results to forecast future performance. Bootstrap technique has some certain advantages. Most importantly, as the distribution of the alphas is complex and non-normal, bootstrap is very convenient to use in the sense that it doesn't require any distribution to be defined. Kosowski et al applied bootstrap techniques to the conditional (as defined by Ferson and Schadt (1996)) and unconditional versions of Carhart's four factor model over the period of 1975 to 2002 on monthly net returns of US open-end domestic equity funds. Although there is certain amount of luck, there is also superior performance, especially in the top 10% of the funds. Another result of the bootstrap tests is that superior fund performance exists for growth oriented funds which is a result first given by Chen, Jegadeesh and Wermers (2000).They also stated that there is no superior fund management ability in income oriented funds. Although there are variations in ability for high alpha funds due to stock picking ability, the differences in low ranked funds are due to expenses rather than skill. Kosowski et al made some checks to assess whether the time series dependence of the residuals have a bad effect on their results. In unreported tests they found that the results are almost identical to their findings with bootstrapping the residuals approach. They also checked whether a possible correlation between the explanatory variables and the residuals have any effect on their results by resampling both the explanatory variables and the residuals. Again they found almost identical results. They finally tested the effect of cross-correlation between fund residuals and found almost identical results. These results suggest that using bootstrapping the residuals approach is a good and simple choice of finding the approximatted distribution of the parameters.

Cuthbertson, Nitzsche and O'Sullivan (2004) applied bootstrap techniques following Kosowski et al (2005) to the largest European mutual fund market, UK. They used 1596 open end mutual funds for the period of April 1975 to December 2002 from which 450 of them are non-surviving to overcome the survivorship bias. They used funds with at least sixty observations to overcome any possible sampling bias. They have classified funds according to their objectives and searched whether outlier performances are based on skill or luck. Cuthbertson et al applied CAPM, multi-factor models of Fama and French (1996) and Carhart (1997), market-timing models of Treynor and Mazuy (1966) and Henriksson and Merton (1981). They have also conditioned these five models following Ferson and Schadt (1996), which conditions only beta to available market information and also Christopherson, Ferson and Glassman (1998), which conditions both alpha and beta to available market information resulting 15 different performance measures of alpha. They have shown that conditioning the models or adding a market timing term do not improve the measure and that the best fit model is the unconditional Fama and French model. Similar to Kosowski et al, Cuthbertson, Nitzsche and O'Sullivan found that there is stock picking ability in best performing funds which cannot be related to chance by itself and that relation is even stronger for worst performing funds. They have measured the performance over the defined classes according to investment objectives and contrary to Kosowski et al, who found stock picking ability for growth oriented funds, they have found strong stock picking ability both for income and growth oriented funds for UK.

4. Data

This study examines monthly returns of 46 mutual funds and 19 pension funds, a total of 65 funds for the period in between January 2004 and December 2009 (their codes and names are available on Appendix A). Mutual funds that are included in this study are A-type, which means at least 25% of the funds for all times are invested in securities issued by Turkish companies. On the other hand pension funds that are included contain at least 15% of their shares as stocks. According to the classification of the Capital Markets Board of Turkey (CMB), 10 of the pension funds that are included in this study are growth oriented, 1 of them is income oriented and 8 of them are classified as “other”. The dataset is obtained from the monthly bulletins of CMB and several corrections are made which are listed in Appendix B.

Pension funds that invest in foreign securities are excluded from the sample, since the benchmark that we use for funds that invest in Turkish securities will not give appropriate results for funds that invest in foreign securities. Similarly, mutual funds with an investment objective of foreign securities are excluded. Moreover, mutual funds that invest in ISE indices are excluded as well, since they are overly correlated with the benchmark and their objective is market tracking (which means they are not trying to outperform the market and so it’s not reasonable to search for performance for these types of funds).

There are 139 funds that meet the above criteria, but are excluded from the sample. There are possible reasons for this exclusion: Funds may merge, the investment companies may go bankrupt, funds may disappear because of poor performance during the observation period or funds may be established during the observation period, which in most cases means they don’t have enough observation. The funds are chosen with the condition that they were born before the observation period and survived until the end of December 2009. This may cause survivorship bias, because the funds that terminated during the observation period are not included. The funds that couldn’t survive until the end of the sample period are the ones which are expected to have poor performance. Excluding these funds from the sample period may cause the performance measure, alpha to be over estimated. The effect of non survivng funds, was analyzed by Blake and Timmermann (1998). They analyzed 2300 UK open-ended mutual funds between February 1972 and June 1995 with the following formula:

$$bias_j = \frac{1}{T_j}\left(r_{s,j,t} - \frac{n_{n,j,t}r_{n,j,t} + n_{s,j,t}r_{s,j,t}}{n_{n,j,t} + n_{s,j,t}}\right) \quad (12)$$

where $n_{n,j,t}$ is the number of non-surviving funds for month t, $n_{s,j,t}$ is the number of surviving funds for month t, $r_{n,j,t}$ is the equally weighted portfolio return of non-surviving funds for month t and $r_{s,j,t}$ is the equally weighted portfolio return of surviving funds for month t.

They have found a considerable amount of survivor bias of 0.8% per annum for UK constructing two portfolios for funds that died during the observation period and for funds surviving until the end of the period. They have further investigated the fund behaviors before termination and after birth and found significant underperformance prior to termination, but slightly over performance in the early periods after birth. They have constructed a zero cost portfolios of best and worst performing funds, monthly rebalanced depending on the past 24 months' average performance and found significant persistence on both sides.

To check whether this bias has an important role for our sample period, survivorship bias measure of Blake and Timmermann (1998) is computed. For the sample period we examine, there are 80 non surviving funds in total and 65 surviving funds that we choose. 0.071% monthly bias corresponds to 0.86% annual bias for our sample period, which at first seems like a similar result for UK, but considering the volatility of Turkish markets, it is less important.

Moreover we examined survivorship bias by regressing the surviving funds' monthly average excess returns on non-surviving funds' monthly average excess returns with the null hypothesis that alpha is zero.

$$R_{s,t} - R_{f,t} = \alpha + \beta\left(R_{n,t} - R_{f,t}\right) \quad (13)$$

This way, if there is any extra performance (which corresponds to a positive and significant alpha) we may conclude that there is survivorship bias for our sample period. From Table (1) we can check that the intercept, alpha is 0.00012 (which is minor) with the probability 0.96. Hence we cannot reject the null hypothesis and we may conclude that our data contains no survivorship bias. Further studies on survivorship bias for Turkish markets are examined at Appendix C.

Table 1: Regression results of surviving funds on non-surviving funds[5]

Dependent Variable: SURVIVING-RF
Method: Least Squares
Sample (adjusted): 2004M02 2009M10
Included observations: 69 after adjustments
Newey-West HAC Standard Errors & Covariance (lag truncation=3)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
NON_SURVIVING-RF	0.791756	0.112055	7.065794	0.0000
C	0.000117	0.002352	0.049862	0.9604

R-squared	0.838327	Mean dependent var	0.001100
Adjusted R-squared	0.835914	S.D. dependent var	0.050955
S.E. of regression	0.020641	Akaike info criterion	-4.894555
Sum squared resid	0.028544	Schwarz criterion	-4.829798
Log likelihood	170.8622	F-statistic	347.4174
Durbin-Watson stat	1.040930	Prob(F-statistic)	0.000000

There is one more bias that we should consider. Our data set consists of funds which have full data during our sample period, but there are 59 funds that were established during the sample period as well as 80 non-surviving funds that we don’t take into account. Not analyzing these 139 funds may cause “exclusion bias” to occur. Regenerating the above regression for monthly excess average returns of included funds and excluded funds gives us an alpha of 0.000571 with probability 0.29. Hence we cannot reject the null hypothesis of zero alpha. Reconstructing Blake and Timmermann (1998)'s measure for exclusion bias gives us the below equation

$$Exc.\,bias_j = \frac{1}{T_j}\left(r_{s,j,t} - \frac{n_{e,j,t}r_{e,j,t} + n_{s,j,t}r_{s,j,t}}{n_{e,j,t} + n_{s,j,t}}\right) \tag{14}$$

where, $n_{e,j,t}$ represents the number of excluded funds for month t and $r_{e,j,t}$ is the equally weighted portfolio return of excluded funds for month t. Our results suggest a 0.32% per annum bias which is negligible.

Table 2: Regression results of included funds on excluded funds

Dependent Variable: SURVIVING-RF
Method: Least Squares
Sample: 2004M02 2009M12
Included observations: 71

Variable	Coefficient	Std. Error	t-Statistic	Prob.
ALL_EXCLUDED-RF	0.937086	0.009977	93.92598	0.0000
C	0.000571	0.000538	1.060632	0.2926

R-squared	0.992239	Mean dependent var	0.001642
Adjusted R-squared	0.992127	S.D. dependent var	0.051116
S.E. of regression	0.004536	Akaike info criterion	-7.925966
Sum squared resid	0.001419	Schwarz criterion	-7.862229
Log likelihood	283.3718	F-statistic	8822.089
Durbin-Watson stat	1.776390	Prob(F-statistic)	0.000000

As a result, our findings suggest a minor bias caused by not taking into account all the funds existing during the period of January 2004 to December 2009. Therefore we may conveniently use the results obtained from the data set.

The benchmark that is used throughout the paper is ISE100 index issued by Istanbul Stock Exchange. Risk free rate is taken to be the 30 days bond index issued by Turkish Institutional Investment Managers Association. Monthly return data is used for fund evaluation. Returns are computed simply as the difference of the price change from prior month to the next one, divided by the price of the prior month.

$$R_{jt} = \frac{P_{jt} - P_{jt-1}}{P_{jt-1}} \tag{15}$$

Where P_{jt} is the price of the fund j at the end of the month t. Prices are published in the monthly bulletin of the Capital Markets Board of Turkey at the end of the month. These prices are announced after deduction of fees and expenses.

5. Methodology

Bootstrap procedure is gaining importance in the literature of fund performance for several reasons. Investigations suggest that varying risk levels of the managers and the non-normalities of the individual fund alphas cause cross section of alpha measures to be non-normal. This contradicts the assumption of normality that many of the performance models used before for checking the statistical validity of the measures. In order to resolve this problem bootstrap technique is used which doesn't impose a distribution for fund performance and allows us to construct a sampling distribution of fund performance. First, we will discuss the generally accepted performance models and apply them to our data set in order to choose the most appropriate models to use in our bootstrap simulations. Then we will argue the advantages of bootstrapping procedure and define the process.

5.1 Regression Models Used

In this section, commonly used performance measures are discussed and appropriate models are selected for intensive bootstrap procedure for Turkish funds. After examining unconditional performance models of Jensen, Fama and French, Carhart, Treynor and Mazuy and Henriksson and Merton for the dataset, we will give beta and alpha and beta conditional results of specific models on Ferson and Schadt and Christopherson, Ferson and Glassman sense, respectively. Since bootstrapping is an intensive process, we will choose best explaining model for each of the three categories (unconditional, conditional on beta and conditional on alpha and beta) and discuss the results. For brevity, reports will be given as the averages of the values estimated. All t statistics are reported as the averages of the absolute values of the t statistics to prevent negative and positive values to cancel each other out, except for the t statistics of the equally weighted portfolios constructed.

5.1.1 Jensen's One Factor Model

The most known performance measure is Jensen's single factor model, which is given by eq. (1). The only risk factor represented in this model is the market. If CAPM is the correct model of equilibrium returns, then the portfolio should be on the Security Market Line (SML), which indicates that the intercept alpha of the regression model should be zero. This implies that, a positive and significant alpha from the regression is the extra performance of the portfolio manager over the market. Jensen claims that if the portfolio manager has some stock picking ability, he will continuously perform better than the market providing this positive alpha. The risk free rate R_{ft} used for Turkish funds as discussed before is the monthly return of the 30 days bond index issued by Turkish Institutional Investment Manager's Association at the end of month t.

$$R_{ft} = \frac{P_{KYD30,t} - P_{KYD30,t-1}}{P_{KYD30,t-1}} \qquad (16)$$

The proxy of the market portfolio which is represented by $\tilde{R}_{Mt}$, is the monthly return of the ISE100 index issued by Istanbul Stock Exchange at the end of month t.

The results of CAPM regression model is presented at the first column of table (3). Average of the alpha estimates of 65 Turkish funds that we consider is -0.22% per month which corresponds to -2.64% annually. This indicates that on average mutual fund managers underperformed the market by an amount of 2.64% per annum. But with 5% significance level, this abnormal performance is insignificant. Although not reported this model gives significant results for only 26% of the funds and alpha performance varies in between the values 1.38% and −1.74% with the best performing fund being ST1 and the worst performing fund being MAD. Models considered after CAPM are more extensive and therefore we expect them to give better estimates of performance.

5.1.2 Treynor and Mazuy's Market Timing Model

Treynor and Mazuy (1966) studied whether portfolio managers try to outguess the market movements and replace their beta, the sensitivity of their portfolio to the market, accordingly. Their assumption of increasing beta as the market moves upwards and decreasing beta as the market moves downwards rejects constant beta. Hence they define beta as a linear function of the market factor, replace it in CAPM regression model and obtain eq. (6).

The results of Treynor and Mazuy timing model are presented in column 3 of table (3). Our regression results on average give 0.21% monthly performance but as CAPM model this performance is not significant at 5%. Negative performance switches to positive performance with Treynor and Mazuy timing model indicating that managers have stock picking ability. Between 2.37% and −1.16% monthly alpha performance results with only 9% of these performances being significant at 5%. Similar to CAPM model, ST1 is the best and MAD is the worst performing fund. The market timing component on average is −49.6% which indicates that Turkish funds cannot time the market although they have stock picking ability. 49% of the funds considered have significant market timing component with 5% significance level. These results indicate that the reason for negative performance coming from CAPM measure is the lack of the timing component. Treynor and Mazuy model corrects this problem.

5.1.3 Henriksson and Merton's Market Timing Model

A similar market timing model is constructed by Henriksson and Merton (1981). They suppose that the manager forecasts whether the market exceeds the risk free rate or not, but don't take in to account the magnitude of it. The manager will have a higher beta if he has a positive forecast and lower beta otherwise. They construct beta as a function of a constant term and a dummy variable (eq. (8)) which gives 1 for positive market forecasts and 0 otherwise. Similar to Treynor and Mazuy, they replace this beta in CAPM regression model and obtain eq. (7).

The results of the Henriksson and Merton timing model are presented in the second column of table (3). We obtain 0.65% monthly performance on average but this value is insignificant at 5%. This alpha performance is better compared to the alpha obtained from Treynor and Mazuy timing model, but still for only 17% of the funds we have significant alphas. The best fund is ST1and it has 3.55% monthly performance with the model and the worst one is MAD with a -0.70% monthly performance. The negative coefficient of the dummy variable which is the measure of market timing indicates that Turkish funds on average have negative market timing ability. 45% of the funds have significant coefficients of market timing at 5% significance level. This result is similar to the result obtained from Treynor and Mazuy timing model.

5.1.4 Fama and French's 3 Factor Model

Fama and French (1993), reports that funds may have investment strategies related to firm characteristics and these should be added to the model as explanatory variables in order to rectify alpha from their misleading effects. They improve Jensen's one factor model, CAPM by adding two additional factors, SMB and HML. SMB stands for "small minus big" index which is a proxy for the size effect and HML stands for "high minus low" index which is a proxy for the effects related to the book-to-market equity ratio. For the calculation of these indices, Fama and French first find the median of the market equities of the NYSE securities and separate them into two classes as "small" and "big", S and B, respectively. Market equity for a fund at time t is calculated as the product of the price and the number of shares at that period. They then divide these securities into three different groups relative to their book-to-market equity ratios (BE/ME) in a way that the highest 30% of these securities are in the "high" group, H, the lowest 30% of them are in the "low" group, L and the middle 40% are in the "middle" group. M. They calculate the book equity (BE) as the book value of the equity plus the deferred taxes and the investment tax credit, minus the book value of the preferred stocks[6]. Then BE/ME is simply BE at the fiscal year ending of calendar year $t-1$ divided by ME at the end of December of year $t-1$. If the firms have negative book-to-market equity ratios, they are excluded from the sample of observations. Fama and French (1992a) in their paper found that book-to-market equity ratio proxy better explains the returns then size proxy, which is why the securities are classified into three groups relative to their book-to-market equities and two groups relative to their sizes. Using these groups, the following six portfolios are formed $S/H, S/M, S/L, B/H, B/M$ and B/L. If a stock is small sized and it has high book-to-market equity ratio for a given time period, then it will be in portfolio S/H. Even though BE/ME and BE are calculated at the end of the year $t-1$, these portfolios are formed six month lagged at the end of June of year t, which is realistic considering the fact that some time will pass until the investors obtain these figures. For each portfolio, returns are weighted according to their market equities. Then, SMB and HML are formed accordingly;

$$SMB = \frac{(S/L + S/M + S/H)}{3} - \frac{(B/L + B/M + B/H)}{3} \tag{17}$$

$$HML = \frac{(S/H + B/H)}{2} - \frac{(S/L + B/L)}{2} \tag{18}$$

We adapt Fama and French's formulation while calculating SMB and HML, but we use publicly available book-to-market equity and market equity values for each year end from the website of ISE. Similarly, price data for month ends of every stock trading in ISE are gathered from the website of ISE in order to calculate the monthly returns of the stocks.

Using eq. (4), we regress the 65 funds that we choose and obtain -0.19% monthly performance on average. This negative performance isn't significant at 5% level. The best and the worst performing funds are ST1 and MAD having 1.54% and -1.69%, respectively. Alpha coefficients are significant for 23% of the funds. Compared to CAPM, there is a slight increase in performance. This is the effect of the factors added. The coefficient of SMB is positive (3.64%) and significant for 37% of the funds where as the coefficient of HML is negative (-4.88%) and significant for 26% of the funds. These positive and negative coefficients imply that managers in Turkey prefer small sized stocks with low book to market equities.

5.1.5 Carhart's 4 Factor Model

Carhart's four factor model is an extension of Fama and French's three factor model. Reports of Jagadeesh and Titman (1993) that portfolios buying winning stocks and selling looser stocks return positive earnings in the short run imply the presence of momentum effect. We should keep in mind that return gained by using this zero investment strategy, which is publicly available, isn't an extra performance generated by the portfolio manager. In order to capture this momentum effect, Carhart uses an additional risk factor, one year momentum proxy to measure manager's performance. For Carhart's model, the proxy of one year momentum effect, $PR1YR$ is the return difference between a portfolio with previously high performing stocks and a portfolio with previously low performing stocks. To construct these two portfolios, we first form prior 11 month returns of all the stocks available with the following formulation;

$$R_{j,prior} = \left[\prod_{n=1}^{11}\left(1 + R_{j,(t-n)}\right)\right] - 1 \qquad (19)$$

where $R_{j,(t-n)}$ is the return of the stock j at time $t-n$. Then we rank the stocks using their prior 11 month returns, $R_{j,prior}$ and form two portfolios using simple average returns of the highest 30% of the funds and the lowest 30% of the funds. Adding this new proxy to Fama and French's three factor model, Carhart's four factor model becomes as in eq. (5).

As can be seen from the fifth column of table (3), our regression results suggest -0.19% monthly average performance which is the same value obtained from Fama and French's three factor model. This performance is insignificant at 5% level. The alpha performance measure varies between 1.55% and -1.69% and 23% of these performances are significant. Similar to Fama and French's model, SMB has positive and HML has negative effect. The negative coefficient of the new factor $PR1YR$ is insignificant on average. Actually this coefficient is significant for only 9 of the funds, which is minor. Further comparisons between the models discussed will be given later for model selection for bootstrap analysis.

Table 3: Regression Results of the Unconditional Models

	1	2	3	4	5
Model	CAPM	HM	TM	FF	C
Average of the Regression Coefficients					
α	-0,0022	0,0065	0,0021	-0,0019	-0,0019
$\lvert t(\alpha)\rvert$	1,3861	1,3354	0,9653	1,4412	1,4298
R_M-R_f	0,5053	0,6211	0,5060	0,5207	0,5201
$\lvert t(R_M-R_f)\rvert$	12,8182	8,3132	12,7108	13,1211	13,0398
SMB				0,0364	0,0362
$\lvert t(SMB)\rvert$				1,7266	1,7787
HML				-0,0488	-0,0485
$\lvert t(HML)\rvert$				1,4144	1,4321
PR1YR					-0,0030
$\lvert t(PR1YR)\rvert$					1,1274
DUMMY=D*(R_M-R_f)		-0,2293			
$\lvert t(DUMMY)\rvert$		1,8745			
$(R_M-R_f)^2$			-0,4958		
$\lvert t((R_M-R_f)^2)\rvert$			1,9541		
Rejection of Normality with Jarque-Bera Test (% of funds)	26,2%	23,1%	23,1%	23,1%	23,1%
R^2	0,7230	0,7410	0,7426	0,7467	0,7523
Adj. R^2	0,7189	0,7334	0,7351	0,7354	0,7373
Equally Weighted Portfolio					
α	-0,0023	0,0063	0,0020	-0,0020	-0,0020
$t(\alpha)$	-1,6143	1,9937	0,9766	-1,3960	-1,4023

5.1.6 Ferson and Schadt Conditional Model

The measures discussed until now are unconditional models of performance, i.e., they don't take into account the changing information about the market. If the portfolio manager changes his portfolio composition and risk in accordance with the changing information about the expected returns and risk of the securities contained in the portfolio, the unconditional measures of performance will give biased results. There are possible reasons for the risk of the portfolios to shift through time: Market corrects the over pricing or the under pricing of securities and companies change their financial strategies. Hence even if the manager follows a buy and hold policy, the risk of the portfolio changes over time because of the risk changes of the underlying securities and the weights of these securities change as well in accordance with their values. Moreover, actively managed portfolios betas and weights will change continuously. To purify the performance measure alpha from the effects of these risk shifts Ferson and Schadt (1997) conditioned their beta to a publicly available information set. Their intuition is that if the manager uses publicly available information, it shouldn't be judged as superior performance and should be eliminated from the performance measure alpha. Eq. (10) gives the beta conditional regression model of CAPM. We may generalize this equation for multiple factor regression models by simply conditioning the betas of the risk factors to the publicly available information set Z_t.

$$r_{jt+1} = \alpha_j + \sum_{i=1}^{K} \beta_{ij} F_{it+1} + e_{jt+1}. \tag{20}$$

If we have a K factor regression model with F_{it+1} defined as the i-th risk factor's portfolio at time $t+1$, β_{ij} as the beta coefficient corresponding to this portfolio and r_{jt+1} as the excess return of fund j at that time period $(\tilde{R}_{jt+1} - R_{ft+1})$, we may convert this unconditional model to a conditional model by replacing the beta coefficients of the risk factors with a linear function of them as defined earlier by eq. (9).

$$r_{jt+1} = \alpha_j + \sum_{i=1}^{K} (\beta_{0ij} + \beta_{uij}^{'} z_t) F_{it+1} + e_{jt+1} \tag{21}$$

$$= \alpha_j + \sum_{i=1}^{K} \beta_{0ij} F_{it+1} + \sum_{i=1}^{K} \beta_{uij}^{'} z_t F_{it+1} + e_{jt+1}.$$

If the information set Z_t has L components, then this Ferson and Schadt conditional model has a constant, K factor portfolios and the product of L information variables with K factor portfolios as regressors, which makes a total of $(L + 1)K + 1$ regressors.

The situation is somewhat different for market timing models. Treynor and Mazuy assume managers with the ability to time the market successfully will adjust their portfolios accordingly and hence define beta coefficient as a linear function of the market. This way the quadratic regression model is obtained (eq. (6)) where γ_j captures the market timing ability of the manager. Defining beta as a linear function of the market helps us capture the timing ability of the manager, but to construct the conditional model, we should include the effects of public information set to this beta. Therefore we define beta as the linear function of both the market and the information set and obtain Ferson and Schadt conditional Treynor and Mazuy timing model;

$$r_{jt+1} = \alpha_j + (\beta_{0j} + \gamma_j r_{Mt+1} + \beta_{uj}^{'} z_t) r_{Mt+1} + \tilde{e}_{jt+1} \tag{22}$$

$$= \alpha_j + \beta_{0j} r_{Mt+1} + \gamma_j (r_{Mt+1})^2 + \beta_{uj}^{'} z_t r_{Mt+1} + \tilde{e}_{jt+1}$$

Where γ_j measures the timing ability of the manager and $\beta_{uj}^{'}$ measures the response to the information available. For Henriksson and Merton's timing model, the situation is much the same as the other models. Since we have two betas for increasing and decreasing markets, we should be careful defining beta as a linear function of the information set. After all dummy variable provides these two betas and so should be analyzed together with β_j.

$$r_{jt+1} = \alpha_j + (\beta_{0j} + \gamma_{0j} D + (\beta_{uj}^{'} + \gamma_{uj}^{'} D) z_t) r_{Mt+1} + \tilde{e}_{jt+1} \tag{23}$$

$$= \alpha_j + \beta_{0j} r_{Mt+1} + \gamma_{0j} D r_{Mt+1} + \beta_{uj}^{'} z_t r_{Mt+1} + \gamma_{uj}^{'} D z_t r_{Mt+1} + \tilde{e}_{jt+1}$$

If the manager has no market timing ability, γ_{0j} and $\gamma_{uj}^{'} z_t$ are zero.

While performing Ferson and Schadt conditional regression models for our data set, we use three publicly available information. Our information set constitutes of a dummy variable "JANUARY" which controls the January effect, "KYD30" proxy which is the lagged return of the KYD30 index issued by the Turkish Institutional Investment Managers Association and "INDUSTRY" proxy which is the lagged percent change in the industrial production index. To form z_t, the vector of deviations from the expected values of these information variables are calculated. We define excepted values as the average values of these variables for the time period considered, which is from February 2004 to December 2009.

Table (4) presents the results of beta conditional versions of previously discussed models. Similar to unconditional version of the models, CAPM, Fama and French three factor model and Carhart four factor model give insignificant negative alpha estimates on average. Also the average alpha estimates being equal to each other continues with the conditional versions of Fama and French and Carhart models. The values are less than the values obtained with unconditional versions, which is a result of the market factors that we condition the betas of the models. 31% of the alpha estimates are significant for CAPM model where as it is 25% for Fama and French and Carhart models. Comparing the significance of the estimates of the factors of the conditional model with the unconditional model shows that there is an increase in the number of significant estimates, which supports the idea of checking conditional versions of the models.

Table 4: Regression Results of the Beta Conditional Models

Model	1 CAPM_FS	2 HM_FS	3 TM_FS	4 FF_FS	5 C_FS
Average of the Regression Coefficients					
α	-0,0027	0,0076	0,0028	-0,0025	-0,0025
|t(α)|	1,5078	1,4904	1,1629	1,4002	1,5225
R_M-R_f	0,4898	0,6440	0,4929	0,5043	0,5163
|t(R_M-R_f)|	13,2635	9,3085	13,1388	13,5167	14,0830
SMB				0,0352	0,0388
|t(SMB)|				1,6253	1,8667
HML				-0,0534	-0,0595
|t(HML)|				1,8152	2,0277
PR1YR					0,0542
|t(PR1YR)|					1,4948
DUMMY=D*(R_M-R_f)		-0,2839			
|t(DUMMY)|		2,2950			
$(R_M-R_f)^2$			-0,7244		
|t($(R_M-R_f)^2$)|			3,0435		
(R_M-R_f)*JANUARY	0,0872	0,1279	0,0257	0,2098	0,2343
|t(R_M-R_f*JANUARY)|	1,0403	1,7590	0,8904	3,3414	3,0173
(R_M-R_f)*KYD30	-10,7011	-42,0701	-24,4760	-6,3619	-9,3421
|t(R_M-R_f*KYD30)|	1,3260	2,3721	2,2493	1,1398	1,2393
(R_M-R_f)*INDUSTRY	0,1574	0,9437	0,4056	0,1872	0,1180
|t(R_M-R_f*INDUSTRY)|	0,9234	2,0911	1,2531	0,7999	0,7487
DUMMY*JANUARY		-0,1458			
|t(DUMMY*JANUARY)|		1,6098			
DUMMY*KYD30		24,9513			
|t(DUMMY*KYD30)|		1,4442			
DUMMY*INDUSTRY		-0,8583			
|t(DUMMY*INDUSTRY)|		1,4008			
SMB*JANUARY				0,1559	0,2339
|t(SMB*JANUARY)|				2,3361	2,3004
SMB*KYD30				-3,2411	-5,6818
|t(SMB*KYD30)|				0,8144	0,9206
SMB*INDUSTRY				0,0130	-0,0249
|t(SMB*INDUSTRY)|				0,9437	0,9607
HML*JANUARY				0,0182	-0,0120
|t(HML*JANUARY)|				1,4514	2,1426
HML*KYD30				12,4863	15,6393
|t(HML*KYD30)|				1,2658	1,6231

Table 4 (Continued)

HML*INDUSTRY				-0,4854	-0,4814
|t(HML*INDUSTRY)|				1,1851	1,2566
PR1YR*JANUARY					0,1047
|t(PR1YR*JANUARY)|					1,9754
PR1YR*KYD30					-9,2807
|t(PR1YR*KYD30)|					1,0001
PR1YR*INDUSTRY					-0,3553
|t(PR1YR*INDUSTRY)|					0,8390
R^2	0,7401	0,7822	0,7645	0,7918	0,8059
Adj. R^2	0,7243	0,7540	0,7464	0,7487	0,7484
Rejection of Normality with Jarque-Bera Test (% of funds)	15,4%	20,0%	21,5%	23,1%	21,5%
Equally Weighted Portfolio					
α	-0,0027	0,0074	0,0027	-0,0025	-0,0025
t(α)	-2,0976	2,3291	1,4968	-1,6789	-1,9528

The results are persistent with the previous findings, that factor models are better than CAPM in explaining performance and that PR1YR component of Carhart's model doesn't help much in improving the explanation power of Fama and French's three factor model, which can be explained by the slight decrease in the adjusted R^2 value.

Alpha estimates of the conditional versions of the timing models are positive and slightly greater than the unconditional versions although insignificant on average with 5% significance. The significance of the measures of timing ability for both models increase compared to the unconditional version of these models especially for Treynor and Mazuy model (71% of the funds have significant estimate of timing ability).

5.1.7 Christopherson, Ferson and Glassman Conditional Model

In a case where the portfolio manager uses more publicly available information than the information set Z_t considered in Ferson and Schadt's conditional model, it may seem like he creates abnormal performance. To eliminate this misleading performance Christopherson, Ferson and Glassman further modify Ferson and Schadt's conditional model. With the idea that this abnormal performance is caused only by the differences in the information sets considered by the model and the manager, they define the performance measure alpha as a linear function of the information set Z_t. This way they try to explain the public information used by the manager which is unknown to us with the public information that we choose to use in the model. Conditioning alpha reduces the time varying abnormal performance and and we obtain Christoperson, Ferson and Glassman's conditional model. Eq. (11) is the simple case of this model for CAPM. To generalize it for other models with more risk factors included as in eq. (20), we simply extend the conditional beta versions of these models by applying conditional alpha approach.

$$r_{jt+1} = \alpha_{0j} + A_j^{'} z_t + \sum_{i=1}^{K} \beta_{0ij} F_{it+1} + \sum_{i=1}^{K} \beta_{uij}^{'} z_t F_{it+1} + e_{jt+1} \tag{24}$$

This alpha and beta conditional model controls investment strategies used by managers to adjust their betas and change their portfolio holdings conditional on the information. Reconstructing the unconditional models discussed above with Chrispotherson, Ferson and Glassman's approach, we provide that Fama and French's model gives the best adjusted R^2 of 0.7488.

The results of the alpha and beta conditional models are presented in table (5). Timing models return positive estimated alphas while factor models return negative estimates aligned with the unconditional and beta conditional versions. All 5 models' average alpha estimates are insignificant at 5% level.

Table 5: Regression Results of the Alpha and Beta Conditional Models

Model	1 CAPM_CFG	2 HM_CFG	3 TM_CFG	4 FF_CFG	5 C_CFG
Average of the Regression Coefficients					
α	-0,0020	0,0073	0,0028	-0,0027	-0,0031
|t(α)|	1,3108	1,5361	1,1431	1,4577	1,6066
R_M-R_f	0,4843	0,6452	0,4892	0,5006	0,5090
|t(R_M-R_f)|	13,5170	9,9633	13,7109	14,0689	14,0402
SMB				0,0279	0,0275
|t(SMB)|				1,3454	1,3494
HML				-0,0653	-0,0692
|t(HML)|				1,9179	2,1136
PR1YR					0,0272
|t(PR1YR)|					1,1847
DUMMY=D*(R_M-R_f)		-0,2733			
|t(DUMMY)|		2,3421			
(R_M-R_f)^2			-0,6658		
|t((R_M-R_f)^2)|			2,8538		
(R_M-R_f)*JANUARY	0,1293	0,1713	0,0599	0,2132	0,3338
|t(R_M-R_f*JANUARY)|	1,7933	1,8008	1,3034	3,2372	2,4173
(R_M-R_f)*KYD30	-11,6585	-56,8984	-24,2162	-8,6779	-11,9135
|t(R_M-R_f*KYD30)|	1,3186	2,4865	2,1750	1,0890	1,2221
(R_M-R_f)*INDUSTRY	0,1933	0,5342	0,3883	0,0717	-0,2575
|t(R_M-R_f*INDUSTRY)|	1,0941	1,3339	1,2418	0,8680	0,8302
JANUARY	-0,0102	0,0120	-0,0055	-0,0026	0,0034
|t(JANUARY)|	1,1831	1,2528	1,1663	1,0136	0,8804
KYD30	-0,2163	-2,0436	-0,1476	0,0972	0,0378
|t(KYD30)|	0,8093	1,7066	0,8636	0,9056	1,0152
INDUSTRY	-0,0570	-0,0492	-0,0418	-0,0581	-0,0679
|t(INDUSTRY)|	1,6620	1,1631	1,2498	1,7139	1,8057
DUMMY*JANUARY		-0,3058			
|t(DUMMY*JANUARY)|		1,5880			
DUMMY*KYD30		59,9972			
|t(DUMMY*KYD30)|		1,8743			
DUMMY*INDUSTRY		-0,0318			
|t(DUMMY*INDUSTRY)|		1,2379			
SMB*JANUARY				0,1133	0,4106
|t(SMB*JANUARY)|				1,4222	1,6644
SMB*KYD30				-1,6623	-5,2992

Table 5 (Continued)

|t(SMB*KYD30)|				0,7612	0,8590
SMB*INDUSTRY				0,0187	-0,1051
|t(SMB*INDUSTRY)|				1,0423	1,0400
HML*JANUARY				0,0094	-0,1134
|t(HML*JANUARY)|				1,3321	1,5054
HML*KYD30				16,0754	22,8541
|t(HML*KYD30)|				1,4659	1,8459
HML*INDUSTRY				-0,5258	-0,2417
|t(HML*INDUSTRY)|				1,1075	0,6909
PR1YR*JANUARY					0,4633
|t(PR1YR*JANUARY)|					1,9057
PR1YR*KYD30					-20,0686
|t(PR1YR*KYD30)|					1,1907
PR1YR*INDUSTRY					-0,9112
|t(PR1YR*INDUSTRY)|					0,8663
R^2	0,7562	0,7950	0,7767	0,8027	0,8172
Adj. R^2	0,7291	0,7568	0,7479	0,7488	0,7491
Rejection of Normality with Jarque-Bera Test (% of funds)	18,5%	26,2%	21,5%	26,2%	26,2%
Equally Weighted Portfolio					
α	-0,0020	0,0070	0,0026	-0,0028	-0,0032
t(α)	-1,2267	2,5544	1,3707	-1,5812	-1,8547

5.1.8 Model Selection

Coefficient of determination, R^2 is a measure to determine how well the model explains future returns. It takes values between 0 and 1. If this coefficient is 1, it means that the regression line is a perfect estimate of our dataset. Since we have alternative models to choose from in order to obtain the best possible model for Turkish funds, it is better to compare the average adjusted R^2 values of the models considered. Adjusted R^2 increases only if the new variable added improves the model more than by chance. It can be negative and is always less than or equal to R^2. It simply adjusts the number of explanatory terms in the model. And what we are trying to do is to decide the explanatory variables for our model. We should also check the p value of the f statistics of the model which controls whether it is statistically significant or not. For all of the 15 models considered, p values of the f statistics are 0 with 0.1% significance.

For the unconditional models of CAPM, Treynor and Mazuy, Henriksson and Merton, Fama and French and Carhart, table (3) presents the results. Carhart's four factor model gives the best adjusted R^2 of 0.7373 which is followed by Fama and French's three factor model with adjusted R^2 of 0.7354, Treynor and Mazuy's timing model with adjusted R^2 of 0.7351, Henriksson and Merton's model with adjusted R^2 of 0.7334 and CAPM model with adjusted R^2 of 0.7189. These values are very close to each other maybe except for CAPM model. Since we want comparability between the unconditional and conditional versions of the bootstrap results we should check the adjusted R^2's of the conditional models before giving the decision for our models. From table (4), Ferson and Schadt conditional version of Henriksson and Merton timing model gives the highest adjusted R^2 value, which is 0.7540. It is followed by beta conditional version of Fama and French three factor model with adjusted R^2 of 0.7487. For alpha and beta conditional versions of these models Henriksson and Merton timing model with adjusted R^2 of 0.7568 is the best performance measure, which is presented in table (5). Carhart's four factor model follows it with adjusted

R^2 of 0.7491 and Fama and French's three factor model with adjusted R^2 of 0.7488. These two values are very close to each other which support the idea that the $PR1YR$ risk factor doesn't add much to explaining the model.

There are two important effects that are masking the true stock picking ability measured by alpha. First of all there might be an under performance in alpha due to a missing timing component. Secondly the alpha might be higher or lower due to passive investment strategies. For bootstrap procedure we are trying to choose one timing and one multi factor model in order to decompose the effect of both the timing inability and the passive investment strategies from the alpha. For the two timing models that we consider Henriksson and Merton's model has a better explanation power for both conditional and unconditional versions. For the other 3 models even though Carhart's four factor model seems like a better choice than Fama and French's three factor model for both unconditional and alpha and beta conditional versions, this very small difference isn't worth considering. Hence we choose Fama and French's three factor model as well as Henriksson and Merton's timing model for our bootstrap procedure for unconditional, beta conditional and alpha and beta conditional versions.

5.2 Bootstrap

Bootstrapping gained importance after the improvements of computer technology since it requires extensive computing power. It is one of the resampling techniques used for making inferences about a given dataset. The key point which separates bootstrap from other well known methods is that it doesn't impose a distribution but rather constructs an empirical one. Traditional tests such as z-test or t-test require an underlying distribution, but bootstrap relies on the information obtained from the known data. This idea to depend on the resource available is given by the phrase "pull yourself up by your own bootstraps" which is behind the name "bootstrap" as well. Traditional tests have strong assumptions and they require mathematical formulas to make inferences using the distribution. But bootstrap not only makes no assumptions about the distribution but also saves us from the mathematical complexity by estimating the empirical distribution with repeated computations.

Bootstrap has little assumptions regarding the data. Since the distribution required is estimated from the dataset at hand, this sample should be a good representative of the population. Another assumption that it has is about the resamples. Considering the fact that bootstrap is a sampling with replacement from a sample, it is assumed that each resample is independent and identically distributed. So bootstrap assumes that the resamples have the same probability distribution, but that they are drawn independently from each other.

Bootstrapping is mainly used for estimating the necessary statistics of an estimator, obtaining confidence intervals, determining the standard error for a parameter or testing a hypothesis. It provides a way of obtaining these results without setting assumptions. This generality makes it applicable to many other problems such as time series analysis, nonlinear regression, survival and reliability analysis and density estimation and many fields of studies such as biology, psychology, physics, genetics, medicine, chemistry, geology, engineering as well as econometrics and accounting.

As Efron (1998) indicated, Fisher's work on maximum likelihood estimation in 1920s influenced bootstrapping. Both bootstrapping and Fisher information use "plug-in principle" which is the substitution of the estimates in return for the unknown parameters. The methods use plug-in principle in a reversed order which is, Fisher information first computes a formula for the necessary statistic and then plugs in the estimates while bootstrap procedure first plugs in the estimates and then uses Monte Carlo simulation to compute the statistic. The important idea is the use of plug-in principle to make inferences.

Other than the early works of Fisher, there are various resampling methods that predate bootstrapping. We will discuss these methods in the next section.

5.2.1 Related Resampling Methods

Although bootstrapping is a popular resampling method, there are some other alternatives. One of these techniques is known as jackknifing which is developed by Quenouille (1949) and further extended by Tukey (1958). In jackknife methodology subsamples are created from the sample by dropping one observation at a time and then averaging the calculated statistic for every subsample. The steps taken for jackknifing a certain statistic θ can be formulated in the following order:

- Observe a sample of size n, $X = \{X_1, \dots, X_n\}$.
- Compute $\hat{\theta}(X)$, the estimator of θ is calculated over the original sample X.
- For $i = 1, \dots, n$, generate a jackknife sample $X^{-i} = \{X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n\}$ by dropping out ith observation from the original sample X.
- For every subsample X^{-i} calculate the relevant statistic $\hat{\theta}_{-i}$.
- Then $\hat{\theta}_* = \frac{1}{n}\sum_{i=1}^{n} \hat{\theta}_{-i}$ gives the jackknifed estimate.
- $\left\{\frac{n-1}{n}\sum_{i=1}^{n}\left(\hat{\theta}_{-i} - \hat{\theta}_*\right)^2\right\}^{1/2}$ gives the jackknife estimate of standard error.
- The estimate of bias is $(n-1)\left(\hat{\theta}_* - \hat{\theta}\right)$.
- So we get the bias corrected jackknife estimate $\tilde{\theta}$ of θ, $\tilde{\theta} = \hat{\theta} - (n-1)\left(\hat{\theta}_* - \hat{\theta}\right) = n\hat{\theta} - (n-1)\hat{\theta}_*$.

Both bootstrapping and jackknifing are similar tools for estimating the bias and the standard error in a statistic without any parametric assumptions. Although they are similar techniques, the procedure they follow is slightly different. In bootstrapping, sampling with replacement is used, but in jackknife procedure subsamples with one less observation than the original sample are formed. If the sample size isn't big (less than 100 or 200), jackknifing is easier to apply than bootstrapping. But jackknife procedure generates only n samples so the information it gives about the statistic is limited. It is a linear approximation to bootstrapping. Therefore, if the estimate of the statistic is nonlinear, jackknifing will give biased results. So bootstrapping is a more general technique than jackknifing. Jackknifing is used mainly for obtaining inferences about cases of complex sampling.

Another resampling method in the literature is cross-validation. It is used to obtain the best fitted model possible for a given dataset. There are several types of cross-validation such as random subsampling validation, k-fold cross-validation and "leave one out" cross-validation. In random subsampling we repeatedly take two subsets randomly from the data, one for training and one for validation. For each such split, models are fit to the training set and then the accuracy of them is tested on the validation set. Then average of the results is taken. Random choise of the splits may result some observations to occur more than once and some not at all, which cause the validation subsamples to overlap. It also causes the results to vary for different random choices. K-fold cross-validation is superior in this sense.

For k-fold cross-validation we divide the data into almost equal k subsets. Set the i-th subsample (i=1,…,k) as the validation set, fit the model to the remaining part and calculate the prediction error of the fitted model for the i-th subsample, which is the expected squared difference between the future realised value and prediction derived from the model. Then make k iterations (so that each subsample becomes a validation set exactly once) of this process and average the estimates of prediction error. If the number k is large, the training sets are large and similar to each other which results cross-validation estimates to have lower bias with higher variance and if the number is small, you have small and less similar training sets resulting higher bias and lower variance in the estimates. If k equals to the number of observations, then for each iteration, only one observation is left out for the validation set and this is called “leave one out” cross-validation. It is usually confused with jackknifing since both techniques omit one observation at a time and work with the remaining data. The distinction between these two procedures is that cross-validation is used to estimate the prediction error and decide on the best fitted model while jackknifing is used to estimate the bias of a statistic.

We know that bootstrapping is more general than jackknifing, but what about the relationship between bootstrapping and cross-validation? Although the methods are asymptotically the same, for small samples cross-validation gives nearly unbiased estimates while simple bootstrapping gives results that are downward biased. Also the results obtained by simple bootstrapping have little variability while it is the opposite for cross-validation. There are other bootstrap methods which are refinements of the simple case, but the downward bias is still an issue for them. Efron and Tibshirani (1995) reports that $.632+$ bootstrap procedure outperforms cross-validation in a catalog of 24 experiments.

Another alternative is the permutation test. It is commonly used for testing the equality of the distributions of two samples. For this purpose first, the two samples are pooled, than from this pool, all permutations of two data sets with the same number of observations as the original samples are generated. In the next step, the test statistic (for example, the difference of means) is calculated for every permutation. P value is simply the number of permutations where the test statistic (difference of means) is higher than the original statistic (difference) divided by the total number of permutations.

Permutation test is a more powerful equality test compared to classical distribution based tests (like z or t tests) because it doesn’t require any mathematical assumption about the distribution. But permutation test wasn’t a popular tool compared to its distribution based counterparts of z and t tests due to the huge amount of computations it requires and the limited capacity of computers in the late 1920s but this isn’t a problem in this century with the technological developments.

Bootstrap resampling is done with replacement, on the other hand permutation tests are resampled without replacement. The permutation tests can be applied to a very limited number of cases compared to bootstrap. That is, the permutation test is about the equality of two sample distributions. So it requires two samples and can only test the equality of the distributions. For example if you want to test the equality of the means of two samples, the variances of them do not need to be equal. This makes the null hypothesis of permutation test unrealistic for some cases.But bootstrapping doesn’t have this kind of limitations so can be applied more generally. Bootstrap standard deviation is a good estimate of the standard error but we cannot say the same thing for permutation test. So when creating a confidence interval, bootstrapping is used instead of permutation resampling.

In order to understand bootstrapping, we must first understand sampling with replacement. In the next section this process is defined with an example.

5.2.2 Sampling with Replacement

In the previous section we discussed several resampling methods and stressed out that bootstrap resampling is done with replacement. In order to understand sampling with replacement, suppose you have five balls named a, b, c, d and e in a basket. From these five balls draw one randomly and record the name. Than make sure you put back this ball in the basket before making another random draw. This process is called sampling with replacement. Assume you picked ball c for your first draw. For your second draw you still have five choices since you put back ball c in the basket before the second draw. Suppose you draw ball e for your second draw. For the third draw, out of the five balls, assume you draw ball e as well. For the next draw you'll have five balls in the basket again. Therefore your record may look like c, e, e, d, a. Since you draw with replacement, bootstrap sampling may contain repeated values or you may have bootstrap resamples that doesn't contain a given observation at all. For the given resample, balls a, c and d appeared only once, but while ball b never appeared, ball e appeared twice. Now we're ready to learn how to bootstrap and different types of it.

5.2.3 Understanding Bootstrap

The idea of bootstrapping is to make inferences about a population characteristic using the dataset at hand. To do so, this method estimates the distribution of the population from the sample with generating large number of iterations. So it is a computer based non-parametric approach. The importantance of bootstrap comes from the fact that it doesn't make any distributional assumptions. Parametric methods not only have strong distributional assumptions but they also require mathematical formulas for calculating the statistics of that distribution. So if the statistic needed isn't distributed as assumed with the model, the results will be invalid with the parametric test. There is also the problem that, in making inferences with parametric tests the estimates of the statistics are needed to be calculated (such as the standard deviation and the mean), but some statistics do not have specific formulas (such as the difference between the medians). Therefore if the distribution of the population is unknown or complicated, it may be better to use the sample at hand and make inferences using the bootstrap procedure instead of making false assumptions about the distribution of the population and using parametric approach.

Bootstrap uses the sample data at hand as if it is the entire population. Then by using Monte Carlo simulation builds the sampling distribution of the statistic. Monte Carlo simulation draws samples from the population randomly, calculates the statistic for each sample and builds the distribution of the statistic. While calculating the statistic, you need to know the distribution of the population. Bootstrap relaxes this assumption since it treats the sample as the entire population and uses resampling with replacement. As discussed in the previous section sampling with replacement offers us data that is slightly different from the original sample and it is done randomly and independently. So the statistics obtained from the resamples will vary slightly. Therefore the distribution obtained from these statistics will be an estimate of the sampling distribution.

To understand how bootstrap works, let's consider a sample with 30 observations generated from a standard normal distribution given in the second column of table (6). Using sampling with replacement 1000 resamples are created. In table (6) first five of the resamples are shown. As a result of sampling with replacement some of the observations may occur more than once and some not at all. As can be seen from the third column of table (6), -0.200 occurs three times in the first resample while 1.191 doesn't occur at all. The last two rows of the table give the mean and the standard deviation of each sample. Since we have 1000 resamples, we have 1000 resample means. The frequency distribution of these values gives us the bootstrap sampling distribution which is shown in figure (1). To check the accuracy of the bootstrap sampling distribution, we checked the normality of this distribution with several tests. The results are given in table (7), which indicates that we cannot reject the null hypothesis of the bootstrapped distribution being normal. We may also check the estimated mean from this distribution which is 0.073 and compare it with the result that we obtained from the parametric distribution of the sample (0.074) to see that they are very similar.

Table 6: Example of Basic Bootstrap Resampling

Number of Observations	Original Sample	Resample 1	Resample 2	Resample 3	Resample 4	Resample 5
1	1.191	-0.200	0.701	-0.210	0.633	1.191
2	0.701	1.358	-0.083	0.413	-2.528	-1.120
3	-0.520	0.153	-0.074	1.222	0.119	1.222
4	-0.074	1.805	-0.200	1.333	-1.193	-0.777
5	-1.226	0.796	1.609	1.609	1.609	1.191
6	1.609	-0.200	1.333	0.969	-2.528	-1.605
7	-2.528	1.222	1.358	1.358	0.796	1.191
8	0.059	1.333	-1.605	-0.198	-0.623	0.119
9	1.805	0.035	-0.623	0.633	1.358	-0.777
10	-0.083	0.153	-1.314	1.507	0.035	1.222
11	0.153	-2.528	0.796	1.358	-0.074	-0.520
12	-0.200	-0.198	0.119	0.633	-0.083	-0.074
13	0.119	-0.210	-1.193	0.969	-0.210	-0.083
14	-0.623	0.059	1.805	0.413	-0.074	1.191
15	-1.314	-1.314	1.358	1.805	-0.200	1.333
16	0.969	0.153	-1.314	1.358	0.633	-0.200
17	0.633	-1.314	-0.200	-1.193	-0.623	1.609
18	1.358	-0.777	0.633	-0.520	0.153	0.413
19	-1.605	0.796	0.119	-0.623	-0.777	0.701
20	0.413	1.805	0.796	1.609	0.701	1.358
21	1.222	-1.314	1.222	0.701	-1.314	-0.777
22	1.507	-0.200	-0.074	1.191	0.153	-0.200
23	-0.777	-1.605	-0.210	1.805	-0.083	-1.605
24	-1.120	-1.226	-1.120	1.191	-1.193	1.191
25	0.796	1.333	0.153	-1.120	-0.777	-0.083
26	-1.193	1.358	-0.210	0.059	0.153	0.035
27	0.035	0.796	0.035	1.191	-0.198	-1.193
28	-0.198	-1.314	0.119	1.609	-0.520	-0.083
29	1.333	0.153	0.119	-1.605	0.796	1.609
30	-0.210	0.413	-0.210	-2.528	-1.226	-1.605
Mean	0.074	0.044	0.128	0.565	-0.236	0.163
St. Dev.	1.074	1.107	0.907	1.113	0.959	1.040

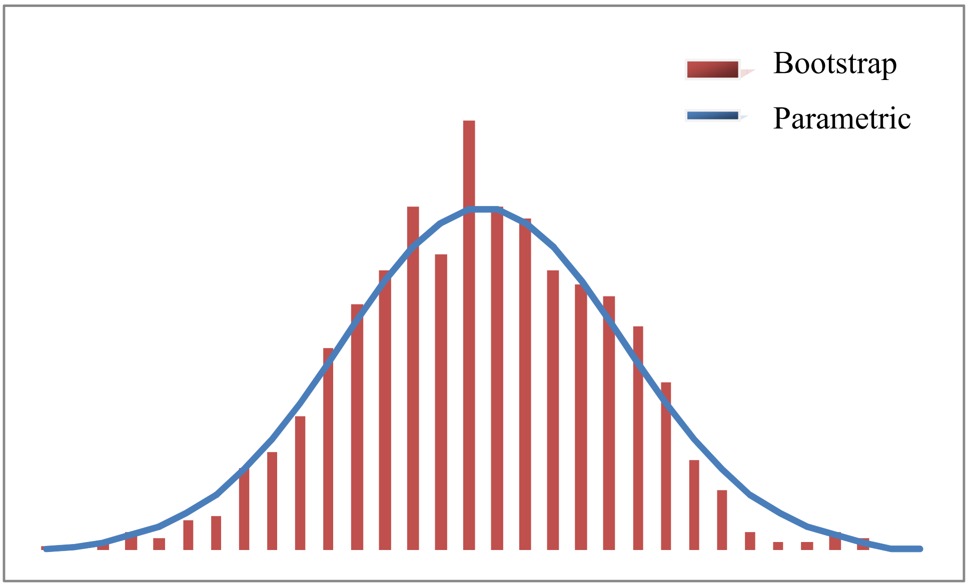

Figure 1: Comparison of Bootstrap and Parametric Sampling Distributions

Table 7: Normality test of the bootstrap means

Method	Value	Adj. Value	Probability
Lilliefors (D)	0.021500	NA	> 0.1
Cramer-von Mises (W2)	0.064366	0.064398	0.3312
Watson (U2)	0.062320	0.062351	0.3141
Anderson-Darling (A2)	0.514734	0.515121	0.1915

5.2.4 Bootstrapping Regression Models

So far we have talked about some resampling methods and gave an example of simple non-parametric bootstrapping to understand the concept more generally. Now bootstrap method is analysed for regression models. Regression models are important tools in statistics which are used to sort out the effects of explanatory variables on a response variable. The inferences that are made about the regression coefficients using parametric methods are based on distributional assumptions. If these assumptions do not hold for the given dataset, then the inferences made may be erroneous. Bootstrap procedure helps to cope with these problems.

Using bootstrap procedure, not only the accuracy of the models can be tested but also hypothesis tests can be constructed, standard errors of the regression coefficients can be estimated and confidence intervals can be formed. Generally there are two bootstrap approaches used for regression models. The first approach for bootstrapping a regression model is to treat the explanatory variables as random and resample the vectors obtained from the observed response variable and the explanatory variables corresponding to that response variable with replacement. It is called case resampling or bootstrapping the pairs (the observations). Suppose we have a standard linear regression model given by $y_i = \sum_{j=1}^{k} x_{ij}\beta_j + \varepsilon_i$ for $i = 1, \dots, n$ with the response variable y and the explanatory variables $x_1, x_2, \dots, x_k$ and number of observations n. Let $z_i = (y_i, x_{i1}, \dots, x_{ik})$ be the i-th vector. For each bootstrap iteration b, resample the vectors z_i and fit the model to obtain bootstrapped regression coefficients $\widehat{\beta}_j^b, j = 1, \dots, k$. Then we can obtain the sampling distribution of the bootstrapped regression coefficients giving equal probability to each coefficient. As explained by Freedman (1981) the error structure of the regression model is ignored with this method. But since resampling is used to imitate the random structure of the sample we should be careful about the choice of the bootstrap methodology.

The second approach for bootstrapping a regression model is to treat the explanatory variables as fixed and resample the residuals. It is called bootstrapping the residuals or error resampling. This method is more complicated compared to the basic non-parametric bootstrap methodology since it first requires a model to be fitted to the sample and the coefficients to be estimated in order to obtain the residuals. So given a sample with n observations, assume the best fitted regression model is of the form $y_i = \sum_{j=1}^{k} x_{ij}\hat{\beta}_j + \hat{\varepsilon}_i$ as in the previous example. After obtaining the regression coefficients, the residual for each observation, $\hat{\varepsilon}_i$ is computed simply by subtracting $\sum_{j=1}^{k} x_{ij}\hat{\beta}_j$ from y_i, i.e, $\hat{\varepsilon}_i = y_i - \sum_{j=1}^{k} x_{ij}\hat{\beta}_j$. For each bootstrap iteration b, resampling procedure is applied to these n residuals randomly with replacement. Then using these residuals $\hat{\varepsilon}_i^b$, the explanatory variables x_{ij} and their estimated coefficients $\widehat{\beta}_j$, bootstrapped response variables are constructed as $y_i^b = \sum_{j=1}^{k} x_{ij}\hat{\beta}_j + \hat{\varepsilon}_i^b$. As indicated before, while computing the bootstrapped response variables and finding the bootstrapped estimates, the structure of the explanatory variables are preserved. To attain the bootstrapped estimates of the explanatory coefficients, the bootstrapped estimates y_i^b are regressed on the explanatory variables x_{ij}. $\hat{\beta}_j^b$'s from the regression model $y_i^b = \sum_{j=1}^{k} x_{ij}\hat{\beta}_j^b + \hat{e}_i$ are the desired coefficients. Again the sampling distribution of these coefficients can be obtained giving equal probability to each bootstrap value. This method assumes that the model selected to obtain the observed residuals is convenient. Therefore if model selection is included in the analysis, the researcher may use case resampling to be on the safe side.

In regression analysis it is common to use ordinary least squares (OLS) estimation which is known to be a very good estimate if the underlying assumptions hold. But if the underlying assumptions of the least squares are violated, then we cannot be sure of the validity of the estimates. The residuals are assumed to be normally distributed while using this approach. Therefore if we assume normality when in reality the error terms are non-normal, the OLS estimates would be misleading. Since the explanatory variables are assumed to be fixed with this estimation technique, any non-normality in the dependent variable may cause the residuals to be non-normal as well, especially in small samples. So we should be careful when the residuals have skewed or heavy-tailed distributions. Also, the outliers may cause error terms to be non-noarmal. Especially in small samples having few extreme values may result a skewed distribution to occur. So in cases like these, bootstrapping procedure provides a method for computing the necessary statistic. The choice of the bootstrap methodology (bootstrapping the residuals or bootstrapping the pairs) depends on the sample of consideration. While bootstrapping the residuals, the residuals are assumed to be independent and identically distributed (iid). Therefore it is ideal to use the other bootstrapping methodology when the residuals are not iid. Since this methodology uses a regression model to find the observed residuals, it is sensitive to the underlying regression model. If the researcher has doubts about the fit of the model, it is best to use bootstrapping the pairs methodology since it is less sensitive to the underlying assumptions. Efron and Tibshirani (1986) claims that the estimate obtained from both techniques are asymptotically equivalent. But one should be careful with small samples. They also emphasize the fact that even if the underlying regression model and the distribution of the residuals do not hold, bootstrapping the residuals can give tolerable results. In some situations, it is better to use bootstrapping the residuals methodology even when the explanatory variables are random. One should decide which method to prefer according to the sample and the explanatory variables available.

5.2.5 Cons and Pros of Bootstrap Sampling

As we discussed in the previous section, bootstrapping can be applied to observations with known distributions (parametric approach) but it is most useful when the distribution of the sample is unknown (non-parametric approach). Bootstrapping emerged as an alternative to traditional statistical methods which assume certain probability distribution. It is generally used to estimate the distribution of a statistic without using normal theory. Therefore it will best work when the distribution is unknown or complicated. Bootstrapping is straightforward to apply, so it has the advantage over analytic methods especially for complex structures. Bootstrapping is a sampling from the empirical distribution of the data, so even for small number of observations; the data will expose its distribution to some tolerable extent. Therefore we may conclude that bootstrapping is useful when the sample size isn’t sufficient for statistical inference. But for very small samples we should be careful to make inferences since even bootstrapping doesn't overcome the weakness of the sample to represent the population, so we cannot be sure of the accuracy of the results. One should also be careful about the number of bootstrap iterations to be sufficient since otherwise there will be a large sampling error. One can overcome this problem by simply increasing the number of iterations to 1000 or more for most of the cases. One shouldn't forget that bootstrapping is an approximation and it is not exact.

5.2.6 Why Do We Use Bootstrap?

As discussed before, differences in the risk taking across funds and non-normalities (skews or fat tails) of individual fund residuals shapes the cross sectional distribution of fund alphas. Considering this complex structure, it may be very difficult to test the significance of the performance measure rationally with an imposed ex ante distribution adapted from the common performance measures discussed earlier in section 5.1. Bootstrap procedure makes no assumptions about the shape of the distribution of the performance measures and hence is a better approach for measuring performance of complex distributions. There is one more fact which makes bootstrap procedure charming. With gigantic markets as in the USA and UK, there is always the question of luck bothering the minds of the investors. Curiosity to find out whether the abnormal performance of a fund is just a consequence of luck or whether it is purely the manager's ability can be fulfilled using bootstrap analysis. Bootstrap analysis helps to detect the impact of luck in performance and figure out the true extent of performance, if any. Even though the size of the Turkish markets is much smaller than these countries and therefore the effect of chance factor may not be as important, it is a growing market and will encounter this problem in near future. So in this study we use bootstrap analysis developed by Kosowski et al (2005) to investigate any stock picking ability beyond chance factor. The procedure can be done using the performance measure alpha or the t statistics of alpha, but t statistics of alpha gives more reliable results considering the fact that it eliminates the effects of risk variability across funds. Since t statistic is estimated dividing by the standard deviation, misleading abnormal performance caused by the heterogeneity of risk taking across funds is eliminated. There are also problems arising from size of the fund and insufficient number of observations which are excluded using t statistics. Markets have a certain depth. For small sized funds, depending on the expectations, it might be easier to sell or buy a higher proportion of the assets of the fund. For example, to increase stock investments by 10% for a fund with net asset value of 100 million TL the fund manager has to buy shares worth 10 million TL. On the other hand, for a fund with total net assets worth 5 million TL, one can manage the same position with only a transaction of 500 thousand TL. So, smaller funds are more mobile and can switch positions easily just because of their size. By using t statistics, this effect is also normalized by the standard deviation in the denominator. If the number of observations is not sufficient, then the average of returns will be sensitive to the outliers. Assume that two funds have an average return of 1% with 99 and 9 observations, respectively. If the next day will be an outlier and both funds return 10%, the effect on average will be 0.1% and 1%, respectively. So having small amount of observations will have alpha distributions with higher variance and have less accurate alpha estimates, which occupy the extreme tails of the cross section of alpha distribution. This misleading performance is also eliminated scaling alpha. Although our dataset excludes funds having little observations and consists of funds that existed during the period February 2004 to December 2009, all having 71 observations, reader should keep in mind their effect for further investigation on fund performance for Turkish markets including these funds.

For the reasons explained above, bootstrap results of t statistics of alpha are evaluated and compared with the observed t statistics of alpha. Newey-West heteroscedasticity and autocorrelation adjusted standard errors are used for derivation of the t statistics. Several performance measures are chosen to obtain healthy and detailed results from the bootstrap analysis. These are the unconditional, beta conditional and alpha and beta conditional versions of Henriksson and Merton timing model and Fama and French three factor model. We discussed these choices on the previous section, section 5.1.8. While using bootstrap analysis for our performance measures we have the null hypothesis of zero abnormal performance ($H_0: \alpha_j = 0$) to determine the luck effect. The procedure is defined in the next section[7].

5.2.7 Bootstrap Procedure for This Study

The bootstrap procedure will be illustrated using unconditional version of Fama and French's three factor model, eq. (4). The application of the procedure to other models is similar and predictable, the only difference being the substitution of the relevant model instead of the unconditional version of Fama and French three factor model.

To prepare for bootstrap procedure our first step is to regress the excess returns of fund j using unconditional version of Fama and French three factor model. For our later study, some of the results of this regression should be stored. These are the coefficient estimates of alpha and factor loadings, $\{\hat{\alpha}_j, \hat{\beta}_{1j}, \hat{\beta}_{2j}, \hat{\beta}_{3j}\}$, t statistics of alpha, $\hat{t}_{\hat{\alpha}_j}$, and time series of estimated residuals, $\{\hat{e}_{jt}: t = T_{j0}, \dots, T_{j1}\}$, where T_{j0} and T_{j1} are the dates of the first and the last monthly returns available for fund j for the time period that we consider. This generalization can be simplified as $\{\hat{e}_{jt}: t = 1, \dots, 71\}$ for this study since for our dataset we only choose funds with available observations for the period February 2004 to December 2009.

For each bootstrap simulation we resample the residuals of fund j stored at step 1 with replacement and obtain the set $\{\hat{e}^b_{jt_\varepsilon}: t_\varepsilon = s^b_{T_{j0}}, \dots, s^b_{T_{j1}}\}$, where b represents the bootstrap iteration number and $s^b_{T_{j0}}, \dots, s^b_{T_{j1}}$ represent the randomly chosen time indices of the resample. For our data set it is simplified as $\{\hat{e}^b_{jt_\varepsilon}: t_\varepsilon = s^b_1, \dots, s^b_{71}\}$. Resampling the residuals with replacement means each time to draw a residual from the basket of residuals randomly, record it and throw it back to the basket before making another draw (Each residual has equal probability of draw). So each sample contains some of the residuals more than once, some of them exactly once and some of them not at all.

For the next step of our procedure we will construct a new time series of monthly excess returns of fund j with the null hypothesis of zero abnormal performance ($\alpha_j = 0$) using the resampled residuals $\hat{e}^b_{jt_\varepsilon}$. Here it is important to note that the factor returns are not resampled.

$$r^b_{jt} = \hat{\beta}_{1j}(r_{Mt}) + \hat{\beta}_{2j}(SMB_t) + \hat{\beta}_{3j}(HML_t) + \hat{e}^b_{jt_\varepsilon} \tag{25}$$

So with the original factor returns and the new resampled residuals we have a new return series r^b_{jt} from the first bootstrap simulation that has a true alpha equal to zero by construction. Reestimating the performance model with this return series may result a positive or negative alpha estimate caused by the choice of the residuals.

Repeating the procedure explained above for all of our funds results a cross section of performance measures to occur for the first bootstrap simulation. We than rank these performance measures which are purely a result of random sampling variability from highest to lowest. Repeating all of the above for all 10000 bootstrap iterations we build distributions of these cross sectional ranked measures with the null hypothesis of zero abnormal performance which are the estimates of random chance. Hence we have a cross sectional distribution of the best performances resulting from 10000 best performances, the distribution of second best performances resulting from 10000 second best performances etc.

Now for a given cross sectional rank, we can compare the actual performance measure with the sampling distribution of the relative rank and decide whether this performance is a result of luck or managerial ability. For example the highest performance measure is compared with the sampling distribution of the highest performance measures with the null hypothesis of zero abnormal performance and we make a decision using the bootstrap p value, which is the probability of observing the actual performance measure simply by chance. It is calculated as the percentage of performance measures of the sampling distribution greater than the actual performance measure.

We may also calculate the number of funds expected to achieve a given level of performance by luck and compare this with the actual results obtained from our data set. As discussed before the performance measure can be either alpha or the t statistics of alpha, but t statistics of alpha gives more reliable results given the fact that it is normalized by standard deviation and it eliminates the misleading results caused by differing risk levels, insufficient number of observations and the size effects. To improve the accuracy of the performance estimates a requirement for the number of observations should be set considering the fact that insufficient number of observations tend to increase the value of the performance estimates. While doing so one should be careful about the survivorship bias (You may need to test the sensitivity of your results with alternative observation requirements). Recent papers ask for a minimum 60 months of observation, but for our data set we don't need this requirement since all funds with no exception have 71 months of observation.

5.2.8 Non-normality of the performance measure alpha

The performance models mentioned above have an assumption of normality for alpha while testing its significance, but we cannot be sure of this normality. Cuthbertson, Nitzsche and O'Sullivan (2004) showed for their data set that around 70% of the mutual funds considered had non-normal residuals. Kosowski et al (2005) shows almost half of the funds (48%) considered in their work have non-normal alpha distributions. There are several reasons for this. First, it is a well known fact that individual share returns have fatter tails compared to normal distribution, resulting non-normalities in fund return distributions. One can argue that a diversified portfolio of non-normal shares will converge to normal distribution, but most fund managers, depending on their expectations, strategies or expertise, usually invest most of their assets into few shares or industries rather than effectively distributing them. Furthermore, the benchmarks followed by the fund managers can exhibit non-normal behavior. In addition, fund returns can exhibit serial correlation. This serial correlation, which is defined as the correlation of a time series with itself, can create patterns and non-normality. Finally, fund managers can shift their risk levels according to some market signals resulting in non-normal individual fund alpha distributions.

One common test for normality is the Jarque-Bera, which is a goodness of fit test with the null hypothesis of normality depending on third and fourth moments. Increasing Jarque-Bera statistic means departure from normality. We have applied the Jarque-Bera normality test to our fund return series. With 5% significance, our test results indicate that 41.5% (27 out of 65 funds) exhibit non-normal distribution.

We have also tested the normality for residuals of performance models discussed in section 5.1. Percentage of funds with non-normal residual distributions varies between 15.4% and 26.2%. For the models that we apply bootstrap, this ratio is slightly higher than the overall average varying between 23.1% and 26.2%. Although the results indicate less non-normally distributed funds compared to Kosowski et al and Cuthbertson, Nitzsche and O'Sullivan's empirical findings, still almost one forth of our funds are non-normal, which is a ratio worth applying bootstrap.

We have discussed the non-normalities of individual fund returns and residuals. Besides these individual non-normalities, the cross section of fund alphas can exhibit non-normal behavior. First reason of this cross sectional non-normality is the non-normality of the individual funds. But, even if the individual funds were normal we still could have observed non-normality in the cross section due to heterogeneous risk taking across funds. Furthermore, due to limited investment choices there might be funds investing in similar portfolios. Cross correlation due to similar investment choices might create fat tails and non-normality. This cross correlation is expected to be more in emerging countries compared to developed countries since the market opportunities are limited.

6. Empirical Results of the Bootstrap Analysis

We have applied bootstrap procedure to our selected 65 funds. The models chosen are Henriksson and Merton timing model and Fama and French three factor model. Henriksson and Merton timing model captures any timing ability. Fama and French three factor model includes two variables called SMB and HML which are used to capture any passive investment strategies. So, in our analysis of alpha performance two models we have chosen can decompose any distortion due to market timing or passive investment strategies. We have also applied the beta and the alpha and beta conditional versions of these models.

The variables used for conditioning for both beta and alpha and beta models are the risk free rate, a January dummy and the industrial production index. The conditional variables are used in the form of deviation from the expected value.

We will present the bootstrap findings in tables (8) through (13). In the order of appearance, results presented are unconditional version Henriksson and Merton timing model, unconditional version of Fama and French three factor model, Ferson and Schadt conditional version of these models as well as Christopherson, Ferson and Glassman conditional version of these models.

We will measure whether the alpha performance is a result of luck or skill. This can be done by using the bootstrapped alphas or t statistics of alphas. As we have discussed earlier, using the bootstrapped alphas can generate misleading results therefore, we will present the results obtained by using the bootstrapped t statistics of alphas. The procedure is simple. Our bootstrapped alpha and t statistics of alpha values are generated through regressing models that are formed with the assumption of no abnormal performance. We calculate the probability of alpha resulting from luck which is the number of bootstrapped t statics of alphas that are greater than the t statistics of actual alpha divided by the total number of bootstrap simulations. If there are 40 bootstrapped t statistics that are greater than our actual t statistic over 10000 bootstrap simulations, than the probability of observing a alpha greater than the actual alpha due to luck is ${}^{40}/_{10000} = 0.4\%$. We call this value the bootstrap p value. P value is the probability of positive skill and in this sense for the worst performing funds 1-p value will be the probability of negative skill.

The results are presented in four columns. First column is the actual alpha itself ranked in a descending order. Second column gives the relative t statistics of these alphas. Third column is the ranked t statistics of the actual alphas. And the fourth column gives the bootstrap p values of the t statistics of column three.

Table 8: Bootstrap Results for Henriksson and Merton Timing Model

	Alpha	T-alpha	T-Stat	P		Alpha	T-alpha	T-Stat	P
1	0,035	3,058	3,412	0	34	0,006	1,295	1,224	0
2	0,027	3,412	3,058	0	35	0,005	1,961	1,205	0
3	0,018	1,657	2,901	0	36	0,005	1,912	1,196	0
4	0,017	2,155	2,861	0	37	0,005	1,193	1,193	0
5	0,016	2,261	2,548	0	38	0,005	0,706	1,171	0
6	0,015	2,265	2,265	0	39	0,004	0,713	1,135	0
7	0,014	2,861	2,261	0	40	0,004	1,041	1,041	0
8	0,014	2,548	2,162	0	41	0,004	0,579	0,998	0
9	0,014	1,611	2,155	0	42	0,004	2,162	0,948	0
10	0,012	1,635	2,071	0	43	0,004	0,685	0,882	0
11	0,011	1,822	2,030	0	44	0,004	1,820	0,868	0
12	0,010	1,299	1,961	0	45	0,003	0,882	0,765	0
13	0,010	1,957	1,957	0	46	0,003	1,171	0,713	0
14	0,010	1,224	1,912	0	47	0,003	1,385	0,706	0
15	0,009	2,030	1,822	0	48	0,003	0,868	0,685	0
16	0,009	0,998	1,820	0	49	0,002	0,446	0,622	0
17	0,009	1,660	1,774	0	50	0,002	0,354	0,616	0
18	0,009	2,901	1,660	0	51	0,002	0,622	0,579	0
19	0,009	1,392	1,657	0	52	0,002	0,765	0,446	0
20	0,008	1,576	1,653	0	53	0,001	0,288	0,384	0
21	0,008	1,599	1,652	0	54	0,001	0,384	0,354	0
22	0,008	1,652	1,635	0	55	0,001	0,616	0,288	0
23	0,008	1,516	1,611	0	56	0,001	0,263	0,263	0
24	0,008	1,531	1,599	0	57	0,000	0,107	0,107	0
25	0,007	1,196	1,579	0	58	0,000	0,107	0,107	0
26	0,007	1,429	1,576	0	59	0,000	0,006	0,006	0
27	0,007	1,774	1,531	0	60	0,000	-0,001	-0,001	0
28	0,007	1,653	1,516	0	61	-0,001	-0,358	-0,358	0
29	0,006	0,948	1,429	0	62	-0,003	-1,036	-0,930	0
30	0,006	1,579	1,392	0	63	-0,003	-1,159	-1,036	0
31	0,006	1,135	1,385	0	64	-0,004	-1,376	-1,159	0
32	0,006	2,071	1,299	0	65	-0,007	-0,930	-1,376	0
33	0,006	1,205	1,295	0					

Henriksson and Merton timing model bootstrap simulation results are presented in table (8). Best performing fund outperforms the market by 3.5% monthly. The t statistic of the best performing fund is 3.058. However when the t statistics are ranked, best t value is 3.412. The bootstrap p value corresponding to this t statistics is zero which means none of the estimated bootstrap t statistics are greater than the actual t statistics of 3.412. P value of zero means that we do not have any evidence of luck even with 1% significance, so we reject the null hypothesis that the actual alpha is a result of luck. Hence we can say that the fund manager has stock picking ability. The alphas of this model range from 3.5% to −0.7%. Results found for the highest ranked t statistics, is consistent with the rest since the p values of all funds are zero without any exception. Hence, according to bootstrap results of Henriksson and Merton timing model, there is significant stock picking ability throughout our dataset.

Fama and French three factor model results show similarities with the Henriksson and Merton timing model. The Fama and French three factor model results are presented in table (9). Best fund alpha is 1.5% monthly, half of the Henriksson and Merton timing model. The t statistics of the best performing fund is 4.520, which is also the highest ranked t statistic. It has a p value of 0, a result similar to Henriksson and Merton timing model, indicating that the alpha achieved is not by chance and rather skill of the manager. As before, bootstrap p values of all funds are zero strengthening our view that the alphas generated are due to skillful managers.

Table 9: Bootstrap Results for Fama and French Three Factor Model

	Alpha	T-alpha	T-Stat	P		Alpha	T-alpha	T-Stat	P
1	0,015	4,520	4,520	0	34	-0,002	-0,813	-0,813	0
2	0,007	2,375	2,375	0	35	-0,002	-1,645	-0,849	0
3	0,006	1,403	1,473	0	36	-0,002	-0,772	-0,889	0
4	0,004	1,473	1,403	0	37	-0,002	-0,756	-0,907	0
5	0,004	1,152	1,358	0	38	-0,002	-0,983	-0,915	0
6	0,004	0,963	1,152	0	39	-0,002	-0,951	-0,951	0
7	0,003	1,358	1,035	0	40	-0,003	-1,509	-0,983	0
8	0,003	1,035	0,963	0	41	-0,003	-1,712	-1,208	0
9	0,002	0,895	0,935	0	42	-0,003	-1,972	-1,447	0
10	0,002	0,935	0,895	0	43	-0,003	-0,907	-1,486	0
11	0,002	0,763	0,834	0	44	-0,003	-1,517	-1,509	0
12	0,001	0,569	0,763	0	45	-0,003	-1,991	-1,517	0
13	0,001	0,834	0,569	0	46	-0,003	-1,208	-1,645	0
14	0,001	0,326	0,326	0	47	-0,004	-1,447	-1,712	0
15	0,001	0,209	0,209	0	48	-0,004	-2,070	-1,753	0
16	0,000	0,191	0,191	0	49	-0,004	-1,486	-1,897	0
17	0,000	0,126	0,126	0	50	-0,004	-0,421	-1,955	0
18	0,000	0,070	0,070	0	51	-0,004	-2,868	-1,972	0
19	0,000	-0,111	-0,098	0	52	-0,005	-1,753	-1,991	0
20	0,000	-0,192	-0,111	0	53	-0,006	-1,955	-2,070	0
21	0,000	-0,132	-0,132	0	54	-0,006	-2,467	-2,115	0
22	-0,001	-0,238	-0,192	0	55	-0,007	-2,481	-2,467	0
23	-0,001	-0,261	-0,238	0	56	-0,007	-1,897	-2,481	0
24	-0,001	-0,098	-0,261	0	57	-0,007	-3,415	-2,613	0
25	-0,001	-0,295	-0,295	0	58	-0,008	-2,613	-2,868	0
26	-0,001	-0,453	-0,421	0	59	-0,008	-4,000	-3,239	0
27	-0,001	-0,473	-0,453	0	60	-0,008	-5,093	-3,415	0
28	-0,001	-0,531	-0,473	0	61	-0,008	-2,115	-3,580	0
29	-0,001	-0,849	-0,531	0	62	-0,008	-4,093	-3,953	0
30	-0,001	-0,682	-0,680	0	63	-0,009	-3,953	-4,000	0
31	-0,002	-0,889	-0,682	0	64	-0,011	-3,239	-4,093	0
32	-0,002	-0,915	-0,756	0	65	-0,017	-3,580	-5,093	0
33	-0,002	-0,680	-0,772	0					

Up until now we have discussed the results of unconditional models. Now we will present the results of the bootstrap simulations of the beta conditional versions. The results of beta conditional Henriksson and Merton model are demonstrated in table (10). Best performing fund, has an alpha of 4.6% per month. Conditioning beta has increased the stock picking ability of the best fund manager. The t statistic of the best fund is 4.225, which is also the highest t statistic available. Similar to unconditional models, p value of the best fund is zero. Alphas for beta conditional Henriksson and Merton timing model ranges in between 4.6% and -0.7%. The bootstrap p values of the t statistics are all zero. This means both good and bad performances are a result of managerial ability.

The other beta conditional model we have analyzed is the Fama and French three factor model. The results of this model are presented in table (11). Compared with the unconditional version of this model, best performing fund alpha decreases by one thirds and becomes 1%. The t statistics associated with this best alpha is 2.545. This number is also the best t statistics of alpha. Bootstrap p value for this best t statistics is 0.0001 which means out of 10000 bootstrap simulation results, only one is greater than 2.545. This number is not enough to assert that chance is the distinctive factor in alpha performance. Alphas range from 1% to -1.7% monthly. All bootstrap p values are zero except for the best t statistics consistent with the results of the previous models. Similar to unconditional models, the beta conditional timing and three factor models strongly support that there is stock picking ability for equity fund managers of Turkey without any exception.

Table 10: Bootstrap Results for Ferson and Schadt conditional Henriksson and Merton Timing Model

	Alpha	T-alpha	T-Stat	P		Alpha	T-alpha	T-Stat	P
1	0,046	4,225	4,225	0	34	0,006	1,595	1,287	0
2	0,031	3,366	4,212	0	35	0,005	1,231	1,231	0
3	0,022	2,662	3,366	0	36	0,005	1,459	1,209	0
4	0,020	2,530	3,308	0	37	0,005	1,502	1,197	0
5	0,019	4,212	2,767	0	38	0,005	1,464	1,114	0
6	0,019	2,435	2,696	0	39	0,004	0,831	1,106	0
7	0,018	1,916	2,662	0	40	0,004	2,049	1,084	0
8	0,017	2,273	2,558	0	41	0,004	0,761	1,082	0
9	0,015	2,767	2,530	0	42	0,004	1,287	1,013	0
10	0,014	1,696	2,435	0	43	0,004	0,725	1,013	0
11	0,013	2,696	2,310	0	44	0,004	0,457	0,831	0
12	0,012	2,198	2,273	0	45	0,003	0,553	0,761	0
13	0,012	1,974	2,218	0	46	0,003	1,409	0,736	0
14	0,012	2,310	2,198	0	47	0,003	0,717	0,725	0
15	0,011	2,218	2,049	0	48	0,003	1,605	0,717	0
16	0,011	1,106	1,974	0	49	0,002	0,486	0,655	0
17	0,011	2,558	1,916	0	50	0,002	0,653	0,653	0
18	0,011	1,715	1,915	0	51	0,002	0,504	0,577	0
19	0,010	1,839	1,849	0	52	0,002	1,114	0,567	0
20	0,010	1,849	1,839	0	53	0,002	0,655	0,553	0
21	0,010	3,308	1,715	0	54	0,002	1,013	0,551	0
22	0,009	1,377	1,696	0	55	0,002	0,736	0,504	0
23	0,008	1,553	1,605	0	56	0,002	0,551	0,486	0
24	0,008	1,464	1,595	0	57	0,002	0,567	0,457	0
25	0,007	1,082	1,553	0	58	0,002	0,275	0,454	0
26	0,007	1,209	1,502	0	59	0,001	0,454	0,275	0
27	0,006	1,197	1,473	0	60	0,000	-0,092	-0,089	0
28	0,006	0,577	1,464	0	61	0,000	-0,089	-0,092	0
29	0,006	1,084	1,464	0	62	-0,002	-0,855	-0,855	0
30	0,006	1,915	1,459	0	63	-0,004	-1,354	-1,120	0
31	0,006	1,473	1,409	0	64	-0,004	-1,515	-1,354	0
32	0,006	1,013	1,403	0	65	-0,007	-1,120	-1,515	0
33	0,006	1,403	1,377	0					

Table 11: Bootstrap Results for Ferson and Schadt Conditional Fama and French Three Factor Model

	Alpha	T-alpha	T-Stat	P		Alpha	T-alpha	T-Stat	P
1	0,010	2,545	2,545	0,0001	34	-0,003	-0,915	-1,083	0
2	0,008	1,790	2,212	0	35	-0,003	-1,337	-1,117	0
3	0,005	2,212	1,790	0	36	-0,003	-1,006	-1,193	0
4	0,004	1,272	1,272	0	37	-0,003	-1,005	-1,239	0
5	0,004	1,051	1,271	0	38	-0,003	-1,193	-1,264	0
6	0,003	0,639	1,051	0	39	-0,003	-1,801	-1,272	0
7	0,003	0,330	0,639	0	40	-0,003	-1,117	-1,314	0
8	0,002	1,271	0,407	0	41	-0,003	-2,120	-1,319	0
9	0,001	0,407	0,382	0	42	-0,003	-1,521	-1,337	0
10	0,001	0,382	0,330	0	43	-0,004	-1,544	-1,367	0
11	0,001	0,204	0,236	0	44	-0,004	-1,494	-1,483	0
12	0,001	0,236	0,204	0	45	-0,004	-2,241	-1,494	0
13	0,000	0,086	0,086	0	46	-0,004	-1,052	-1,521	0
14	0,000	0,062	0,062	0	47	-0,004	-1,483	-1,544	0
15	0,000	0,022	0,022	0	48	-0,004	-2,480	-1,711	0
16	0,000	-0,023	-0,023	0	49	-0,004	-1,239	-1,770	0
17	0,000	-0,025	-0,025	0	50	-0,005	-1,770	-1,801	0
18	0,000	-0,131	-0,131	0	51	-0,005	-1,367	-1,819	0
19	0,000	-0,132	-0,132	0	52	-0,005	-1,819	-2,120	0
20	0,000	-0,226	-0,226	0	53	-0,006	-2,384	-2,159	0
21	-0,001	-0,663	-0,448	0	54	-0,006	-2,289	-2,241	0
22	-0,001	-0,768	-0,564	0	55	-0,006	-1,711	-2,289	0
23	-0,001	-0,578	-0,578	0	56	-0,007	-2,428	-2,384	0
24	-0,001	-0,652	-0,652	0	57	-0,007	-1,264	-2,428	0
25	-0,002	-0,564	-0,663	0	58	-0,007	-3,285	-2,480	0
26	-0,002	-0,829	-0,768	0	59	-0,009	-4,461	-3,285	0
27	-0,002	-0,829	-0,829	0	60	-0,009	-3,419	-3,312	0
28	-0,002	-0,448	-0,829	0	61	-0,009	-3,383	-3,383	0
29	-0,002	-1,314	-0,915	0	62	-0,009	-2,159	-3,419	0
30	-0,002	-1,013	-1,005	0	63	-0,009	-4,156	-3,882	0
31	-0,002	-1,083	-1,006	0	64	-0,010	-3,882	-4,156	0
32	-0,002	-1,319	-1,013	0	65	-0,017	-3,312	-4,461	0
33	-0,002	-1,272	-1,052	0					

There is various publicly available information that the factors are conditioned on. There might be certain information left outside the information set used to condition the factors that are determined by the fund manager. Christopherson, Ferson and Glassman argue that the missing information may have a certain effect on the performance measure alpha and this effect can be conditioned on the information set included in the model. We have applied this approach developed by Christoperson et al to our performance models and estimated both Henriksson and Merton timing model and Fama and French three factor model in alpha and beta conditional sense. Results for Henriksson and Merton timing model are presented in table (12). Best performing fund's alpha is 5% monthly, slightly higher than both the unconditional and beta conditional versions. The t statistic of the best performing fund is 5.282. It is also the highest t statistic. The alphas range from 5% to -0.7%. The bootstrap p values, including the best fund, are all zero. This indicates that even with conditioned alphas the results do not change that the alphas generated by Turkish fund managers are due to their skill.

Last model we have estimated the bootstrap t statistics is the alpha and beta conditional version of the Fama and French three factor model. The results are demonstrated in table (13). Best alpha performance is 1% monthly and the t statistics related to this value is 2.662. This number is also the best t statistics attained from the regression model. Bootstrap p value corresponding to this t statistic is 0.0001 meaning only one of the bootstrap t statistics exceeded 2.662. For this model, alpha performance varies between 1% and -1.6%, monthly. All bootstrap p values for the t statistics are zero with the exception of the best fund according to t statistics. This means we strongly reject the null hypothesis and conclude that managers are skillful.

Table 12: Bootstrap Results for Christopherson, Ferson and Glassman Conditional Henriksson and Merton Timing Model

	Alpha	T-alpha	T-Stat	P		Alpha	T-alpha	T-Stat	P
1	0,050	5,282	5,282	0	34	0,005	1,287	1,248	0
2	0,031	3,596	4,625	0	35	0,005	2,022	1,248	0
3	0,022	2,457	4,434	0	36	0,005	1,449	1,203	0
4	0,021	3,324	3,633	0	37	0,005	1,688	1,184	0
5	0,021	2,349	3,596	0	38	0,005	0,853	1,079	0
6	0,020	2,953	3,324	0	39	0,004	1,011	1,011	0
7	0,020	4,625	2,984	0	40	0,004	1,256	0,935	0
8	0,018	2,197	2,953	0	41	0,004	1,248	0,888	0
9	0,017	3,633	2,684	0	42	0,003	0,563	0,864	0
10	0,014	2,650	2,650	0	43	0,003	0,647	0,853	0
11	0,014	1,638	2,630	0	44	0,003	0,846	0,846	0
12	0,013	2,984	2,457	0	45	0,003	0,935	0,838	0
13	0,012	2,264	2,349	0	46	0,002	1,184	0,647	0
14	0,011	4,434	2,308	0	47	0,002	0,381	0,563	0
15	0,011	2,223	2,285	0	48	0,002	0,378	0,474	0
16	0,011	2,684	2,264	0	49	0,002	0,864	0,462	0
17	0,010	1,392	2,223	0	50	0,002	0,838	0,381	0
18	0,010	1,496	2,197	0	51	0,001	0,474	0,378	0
19	0,009	2,308	2,022	0	52	0,001	0,102	0,254	0
20	0,009	1,883	1,883	0	53	0,001	0,243	0,247	0
21	0,009	2,285	1,836	0	54	0,001	0,247	0,243	0
22	0,009	1,836	1,688	0	55	0,001	0,254	0,132	0
23	0,008	0,888	1,638	0	56	0,000	0,114	0,114	0
24	0,008	1,470	1,619	0	57	0,000	0,132	0,102	0
25	0,007	1,254	1,559	0	58	0,000	0,032	0,032	0
26	0,007	1,203	1,496	0	59	0,000	-0,010	-0,010	0
27	0,006	2,630	1,470	0	60	-0,001	-0,300	-0,252	0
28	0,006	1,619	1,449	0	61	-0,001	-0,252	-0,300	0
29	0,006	1,251	1,392	0	62	-0,003	-1,003	-1,003	0
30	0,006	1,559	1,287	0	63	-0,004	-1,396	-1,070	0
31	0,006	0,462	1,256	0	64	-0,005	-1,615	-1,396	0
32	0,005	1,248	1,254	0	65	-0,007	-1,070	-1,615	0
33	0,005	1,079	1,251	0					

Table 13: Bootstrap Results for Christopherson, Ferson and Schadt Conditional Fama and French Three Factor Model

	Alpha	T-alpha	T-Stat	P		Alpha	T-alpha	T-Stat	P
1	0,010	2,662	2,662	0,0001	34	-0,003	-1,055	-1,140	0
2	0,009	2,281	2,281	0	35	-0,003	-0,656	-1,141	0
3	0,004	1,909	1,909	0	36	-0,003	-1,645	-1,216	0
4	0,004	1,264	1,264	0	37	-0,003	-1,448	-1,270	0
5	0,004	1,156	1,175	0	38	-0,004	-1,825	-1,329	0
6	0,003	0,363	1,156	0	39	-0,004	-2,176	-1,333	0
7	0,003	0,640	0,640	0	40	-0,004	-1,993	-1,361	0
8	0,003	1,175	0,412	0	41	-0,004	-1,058	-1,395	0
9	0,001	0,412	0,363	0	42	-0,004	-1,140	-1,436	0
10	0,001	0,314	0,314	0	43	-0,004	-1,116	-1,446	0
11	0,001	0,301	0,301	0	44	-0,004	-2,253	-1,448	0
12	0,000	-0,020	-0,020	0	45	-0,004	-1,112	-1,480	0
13	0,000	-0,169	-0,121	0	46	-0,004	-1,436	-1,645	0
14	0,000	-0,121	-0,137	0	47	-0,004	-2,180	-1,651	0
15	0,000	-0,137	-0,160	0	48	-0,004	-1,480	-1,729	0
16	-0,001	-0,160	-0,169	0	49	-0,004	-1,901	-1,825	0
17	-0,001	-0,444	-0,250	0	50	-0,004	-1,395	-1,901	0
18	-0,001	-0,353	-0,353	0	51	-0,005	-1,270	-1,993	0
19	-0,001	-0,250	-0,379	0	52	-0,005	-2,161	-2,069	0
20	-0,001	-0,839	-0,444	0	53	-0,006	-1,651	-2,151	0
21	-0,001	-0,628	-0,554	0	54	-0,006	-2,151	-2,161	0
22	-0,001	-0,936	-0,628	0	55	-0,007	-1,729	-2,176	0
23	-0,002	-0,554	-0,656	0	56	-0,007	-1,329	-2,180	0
24	-0,002	-1,006	-0,731	0	57	-0,007	-3,261	-2,253	0
25	-0,002	-0,379	-0,839	0	58	-0,008	-2,464	-2,464	0
26	-0,002	-0,935	-0,935	0	59	-0,009	-2,069	-3,040	0
27	-0,002	-0,731	-0,936	0	60	-0,009	-3,040	-3,078	0
28	-0,002	-1,216	-1,006	0	61	-0,009	-5,403	-3,261	0
29	-0,002	-1,101	-1,055	0	62	-0,009	-3,325	-3,325	0
30	-0,003	-1,333	-1,058	0	63	-0,010	-4,488	-3,719	0
31	-0,003	-1,141	-1,101	0	64	-0,010	-3,719	-4,488	0
32	-0,003	-1,361	-1,112	0	65	-0,016	-3,078	-5,403	0
33	-0,003	-1,446	-1,116	0					

Results indicate that, all the funds have stock selection skill. Although all versions provide the same result for stock picking ability, level of alphas significantly change between Henriksson and Merton timing model and Fama and French three factor model. The average alpha of Henriksson and Merton timing model for all versions are positive ($0.65\%, 0.76\%$ and 0.73%), while it is negative ($-0.19\%, -0.25\%$ and -0.27%) for all Fama and French three factor models (The average alphas are obtained from tables (3), (4) and (5), respectively). These results indicate that while the alpha is negative for Fama and French, it switches to positive with the addition of timing component. Negative timing and positive stock picking ability is what we obtain from timing models. So, we can say that the underperformance of alpha in Fama and French three factor model is due to the missing timing component. From our bootstrap results we can conclude that there is stock picking ability for Turkish fund managers.

Figure (2) presents histograms of the best, the second best, the fifth and the tenth ranked fund alphas obtained by bootstrap simulations with Henriksson and Merton timing model. The best ranked fund alphas from the bootstrap simulation have a right skewed distribution. This skewness means there are exceptionally high alphas. As we move from the best ranked fund towards the center, the distribution of the bootstrap simulation of the alphas converges to normal. We may also observe from this figure that variance decreases as we move from the best ranked fund towards the center which is predictable since higher ranked funds generate abnormal performance and this increases the variability and the middle ranked funds generate modest performances with small variances. Of course this high variance and non-normality of the best ranked fund alphas is closely related to the distribution of the actual fund residuals.

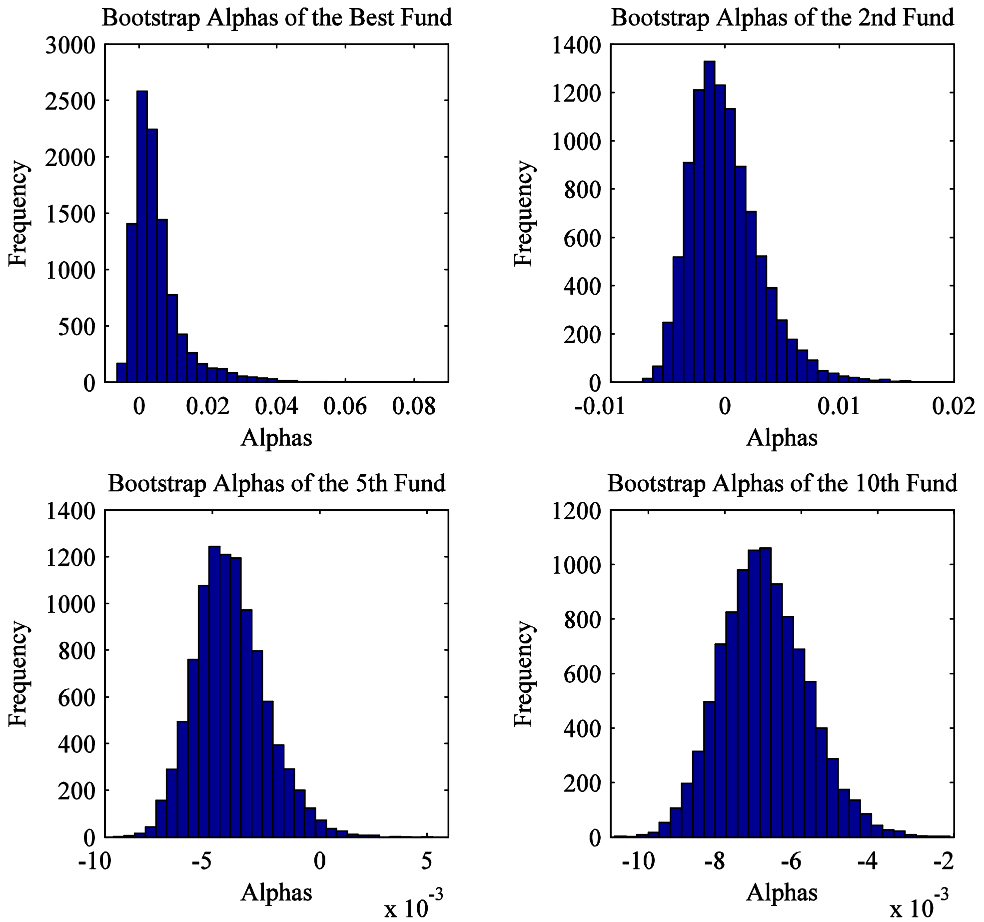

Figure 2: Upper half bootstrap alpha histograms obtained using Henriksson and Merton Model

Figure (3) presents the histogram of the residuals for selected points of the performance distribution. Best ranked fund residuals have higher variances relative to the middle ranked funds. They also have slightly higher values relative to the middle ranked fund residuals. Both these reasonings contribute bootstrap simulation estimates of the alphas to become non-normal for higher ranked funds. The results obtained by Fama and French three factor model are parallel to Henriksson and Merton's model and therefore only one representative figure is shown.

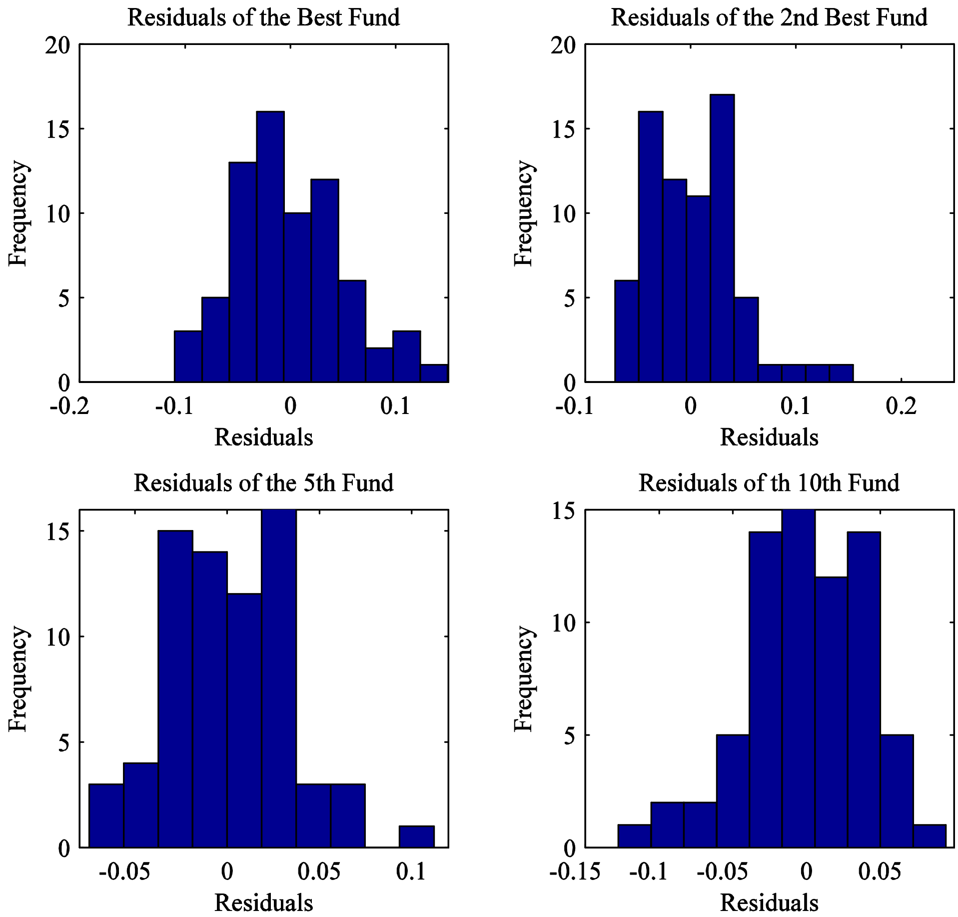

Figure 3: Histograms of the upper ranked alphas' residuals obtained by Henriksson and Merton timing model

Figure (4) and (5) present the histograms of lower ranked fund bootstrap alphas and actual residuals, respectively. These results are obtained using Henriksson and Merton timing model and as before Fama and French three factor model presents results parallel to Henriksson and Merton's. So for brevity only timing model results will be shown. As can be seen from figure (4), worst ranked fund bootstrap alpha histogram is highly non-normal and left skewed. When moved toward the middle ranked funds, the histogram of the bootstrap alphas converges to normal. Variance of the distributions is higher for worst ranked funds and it decreases towards the middle ranked funds. Again, the non-normal distributions of the residuals affect these simulation results. The variance of the worst fund residuals is higher than the variances of the middle ranked funds which cause the non-normality of the bootstrap simulations. These results are mirror images of the upper ranked funds and are expected.

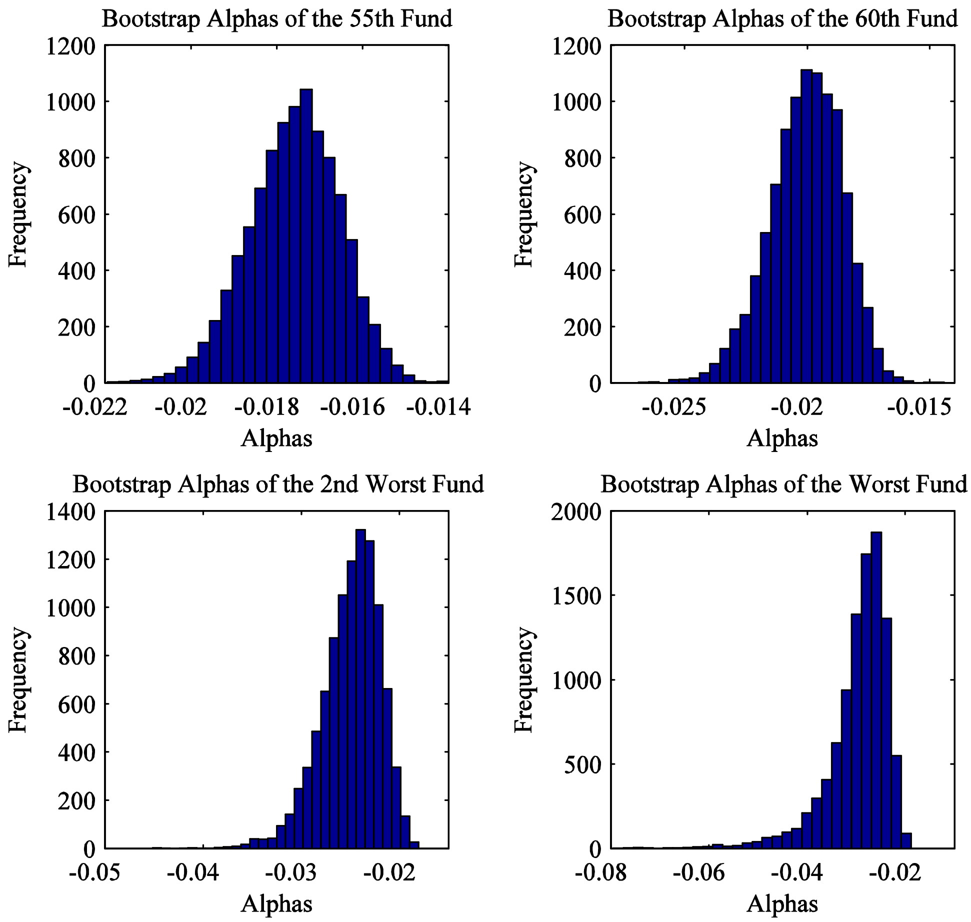

Figure 4: Lower half bootstrap alpha histograms obtained using Henriksson and Merton Model

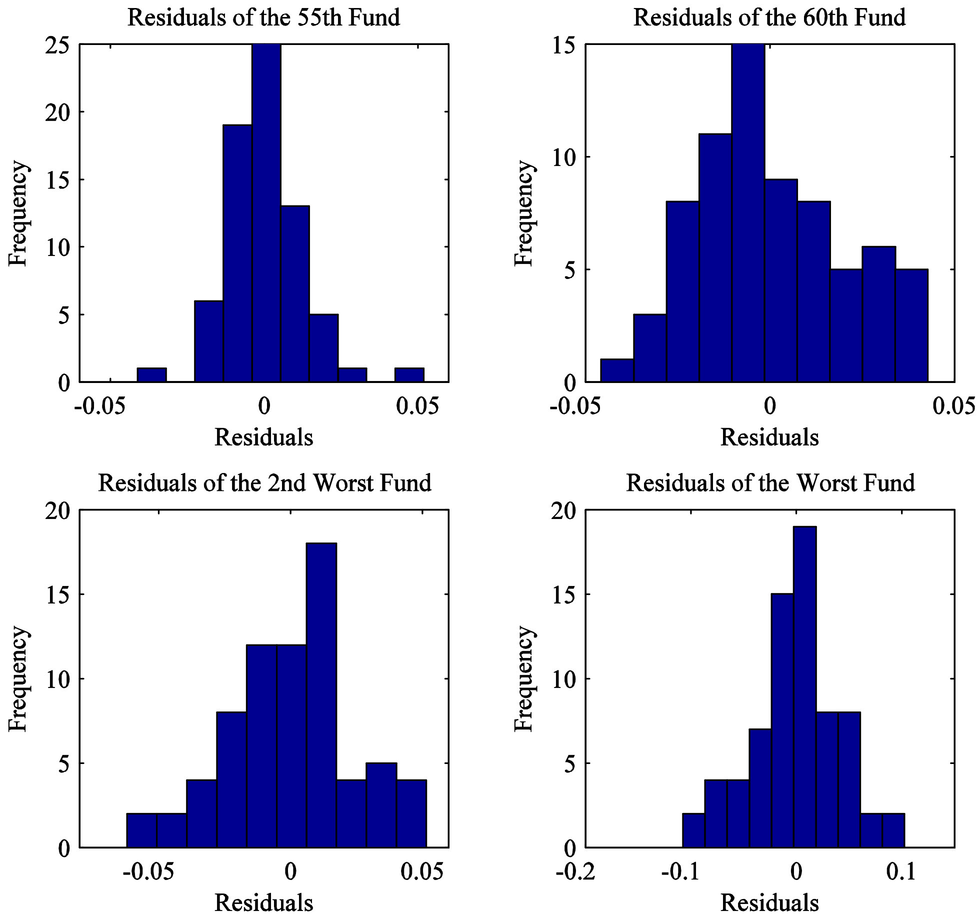

Figure 5: Histograms of the upper ranked alphas' residuals obtained by Henriksson and Merton timing model

7. Conclusion

In this study we have investigated any stock picking ability and whether this ability is due to skill or luck for 65 selected Turkish funds during the period February 2004 to December 2009. Our data contains only funds that existed during the time period that we consider. To make sure of the validity of our results we have checked any bias that may occur due to the exclusion of funds, which are either disappeared or issued during the observation period. For this purpose we have applied two different tests, one of which is Blake and Timmerman's (1998) approach. The calculations showed 0.86% annual bias which is minor for Turkish markets and therefore it may be neglected. For the other test, we have regressed excluded funds to the included funds to check whether any alpha is generated which will correspond to the amount of bias created. The resulting alpha is 0.00012 which verifies that our data is appropriate to interpret economy wide results.

We have applied various regression models to our dataset. We began with CAPM and extended it with two timing models and two multi factor models. Timing models used are well known measures of Treynor and Mazuy and Henriksson and Merton. Multi factor models used are the three factor model of Fama and French and its four factor extension of Carhart which adds momentum factor. We have found that most of the alpha coefficients are negative for multi factor models while they are positive for timing models. Another finding was that most of the timing components of the timing models are negative. So we have concluded that although Turkish fund managers have certain amount of stock picking ability, this ability is buried under the bad timing of the managers.

The non-normality of the distributions of fund alphas as well as the heterogeneous risk taking among funds causes joint distribution of the alphas to be non-normal which violates the normality assumption accepted by the well known performance measures. To overcome this problem, we have adapted bootstrap methodology, which makes no assumption for the distribution of the alphas and constructs an empirical joint distribution instead. We have employed a timing (Henriksson and Merton timing model) and a multi factor (Fama and French three factor model) model in unconditional, beta conditional and alpha and beta conditional versions for the bootstrap simulations. This way the differences between the timing and the multi factor models have been compared. We have found that fund managers in Turkey have stock picking ability but they lack timing.

With the bootstrap procedure we have also searched for any evidence of chance factor in alphas generated. We have bootstrapped the selected models with the null hypothesis of zero abnormal performance and compared the results obtained with the actual measures. Instead of alpha, we have compared the t statistics of alphas in order to overcome the heterogeneous risk levels of the funds. We have calculated a p value which is the probability of receiving a bootstrap t statistic higher than the actual one with the null hypothesis of alpha being zero. While a p value smaller than 5% corresponds to a positive stock picking ability, a p value higher than 95% corresponds to a negative stock picking ability. Values between these percentages mean that the alphas generated are due to luck. Our results for all models are similar and they indicate that the p values are smaller than 5% (even 1%) for all funds indicating that there is stock picking ability for Turkish fund managers. Negative performance cannot be explained by the stock picking inability of the manager.

Bootstrap approach provides an easy solution to estimating the distribution of the desired statistic without getting deep inside the parametric world where there are complex analytical formulas. It gives an approximated distribution when parametric tests fail. The researcher may be interested in using this advantage of bootstrapping to further develop performance analysis in general and check its application to hedge funds. They may also check any evidence of stock picking ability for certain types of funds.

8. Appendix

8.1 Appendix A - Abbreviations of the Funds

65 funds used in this study are listed below.

Code	Name of Fund
AAK	Ata Yatirim Menkul Kiymetler A.Ş. A Tipi Karma Fon
ACD	Acar Yatirim Menkul Değerler A.Ş. A Tipi Değişken Fonu
ADD	Anadolubank A.Ş. A Tipi Değişken Fon
ADF	Akbank T.A.Ş. A Tipi Değişken Fon
AE3	Avivasa Emeklilik Ve Hayat A.Ş. Büyüme Amaçli Esnek Eyf
AH0	Anadolu Hayat Emeklilik A.Ş. Büyüme Amaçli Esnek Eyf
AH5	Anadolu Hayat Emeklilik A.Ş. Büyüme Amaçli Hisse Senedi Eyf
AH9	Anadolu Hayat Emeklilik A.Ş. Esnek Eyf
AK3	Akbank T.A.Ş. A Tipi Hisse Senedi Fonu
AN1	Alternatifbank A.Ş. A Tipi Değişken Yatirim Fonu
ANE	Aegon Emeklilik Ve Hayat A.Ş. Dengeli Eyf
ANS	Aegon Emeklilik Ve Hayat A.Ş. Gelir Amaçli Hisse Senedi Eyf
ASA	Alternatifbank A.Ş. A Tipi Hisse Senedi Fonu
AVD	Avivasa Emeklilik Ve Hayat A.Ş. Dengeli Eyf
AVE	Avivasa Emeklilik Ve Hayat A.Ş. Esnek Eyf
AZB	Allianz H.E.A.Ş. Büyüme Amaçli Esnek Eyf
BEE	Başak Groupama Emeklilik A.Ş. Esnek Emeklilik Yatirim Fonu
BEH	Başak Groupama Emeklilik A.Ş. Büyüme Amaçli Hisse Senedi Eyf
DAH	Denizbank A.Ş. A Tipi Hisse Senedi Fonu
DZA	Denizbank A.Ş. A Tipi Değişken Fon
DZK	Denizbank A.Ş. A Tipi Karma Fon
EC2	Eczacibaşi Menkul Değerler A.Ş. A Tipi Değişken Fon
EV1	Evgin Yatirim Menkul Değerler Ticaret A.Ş. A Tipi Karma Fon
FAF	Finansbank A.Ş. A Tipi Hisse Senedi Fonu
FEE	Fortis Emeklilik Ve Hayat A.Ş. Esnek Eyf
FI2	Finansbank A.Ş. A Tipi Değişken Fon
FYD	Finans Yatirim Menkul Değerler A.Ş. A Tipi Değişken Fonu
FYK	Finans Yatirim Menkul Değerler A.Ş. A Tipi Karma Fon

GAF Gedik Yatirim Menkul Değerler A.Ş. A Tipi Hisse Senedi Fonu
GAK Gedik Yatirim Menkul Değerler A.Ş. A Tipi Karma Fon
GBK Global Menkul Değerler A.Ş. A Tipi Karma Fon
GEH Garanti Emeklilik Ve Hayat A.Ş. Büyüme Amaçli Hisse Senedi Eyf
GL1 Global Menkul Değerler A.Ş. A Tipi Değişken Yatirim Fonu
HLK T.Halk Bankasi A.Ş. A Tipi Karma Fon
HSA Hsbc Bank A.Ş. A Tipi Değişken Fon
IEH Ing Emeklilik A.Ş. Büyüme Amaçli Hisse Senedi Eyf
IEK Ing Emeklilik A.Ş. Büyüme Amaçli Karma Eyf
IGH Ing Bank A.Ş. A Tipi Hisse Senedi Yatirim Fonu
IYD Iş Yatirim Menkul Değerler A.Ş. A Tipi Değişken Fon
MAD Meksa Yatirim Menkul Değerler A.Ş. A Tipi Değişken Fon
SAD Şekerbank T.A.Ş. A Tipi Değişken Fon
SMA01 Sanko Menkul Değerler A.Ş. A Tipi Değişken Yatirim Fonu
ST1 Strateji Menkul Değerler A.Ş. A Tipi Değişken Fon
TAD Taib Yatirim A.Ş. A Tipi Değişken Fonu
TAH Tekstil Bankasi A.Ş. A Tipi Hisse Senedi Fonu
TCD Tacirler Menkul Değerler A.Ş. A Tipi Değişken Fon
TE3 Türk Ekonomi Bankasi A.Ş. A Tipi Karma Fon
TI2 T.Iş Bankasi A.Ş. A Tipi Hisse Senedi Fonu
TI3 T.Iş Bankasi A.Ş. A Tipi Iştirak Fonu
TI7 T.Iş Bankasi A.Ş. A Tipi Değişken Fonu
TKF Tacirler Menkul Değerler A.Ş. A Tipi Karma Fon
TKK T.Iş Bankasi A.Ş. A Tipi Karma Kumbara Fonu
TUD Turkish Yatirim A.Ş. A Tipi Değişken Fonu
TYH Teb Yatirim Menkul Değerler A.Ş. A Tipi Hisse Senedi Fonu
TZD Ziraat Yatirim Menkul Değerler A.Ş. A Tipi Değişken Fon
TZK T.C.Ziraat Bankasi A.Ş. A Tipi Karma Fonu
VAF Türkiye Vakiflar Bankasi T.A.O. A Tipi Değişken Fonu
VEE Vakif Emeklilik A.Ş. Esnek Eyf
VEH Vakif Emeklilik A.Ş. Büyüme Amaçli Hisse Senedi Eyf
YAD Yatirim Finansman Menkul Değerler A.Ş. A Tipi Değişken Fon

YAF Yapi Kredi Yatirim Menkul Değerler A.Ş. A Tipi Değişken Fonu

YAK Yapi Ve Kredi Bankasi A.Ş. A Tipi Karma Fonu

YEE Yapi Kredi Emeklilik A.Ş. Esnek Eyf

BZA Bizim Menkul Değerler A.Ş. A Tipi Hisse Senedi Fonu

YEH Yapi Kredi Emeklilik A.Ş. Büyüme Amaçli Hisse Senedi Eyf

8.2 Appendix B - Data Corrections

The dataset is obtained from the CMB monthly bulletins. This huge dataset may contain certain errors resulting from incautiousness of CMB specialists during the preparation of the bulletins, our incautiousness while consolidating monthly returns or code changes that went unnoticed through time. Therefore we made several corrections. February 2009 bulletin had incorrect prices so we obtained the correct data from the historical fund prices statistics page of CMB. For the funds EV1 and FYD several price data (February, March and June, July, respectively) were missing for 2005, hence we obtained these values from the historical fund prices. June 2008 data for IGH, December 2006 and October 2007 data for TAD, December 2008 data for both TZD and YAD, January 2006 data for SMA and July 2007 data for MAD were missing therefore we obtained these values from the webpage "fonbul[8]". There are also several code changes. GPR was coded as BDY in the monthly bulletins of March, April and May 2004. YKA fund was coded as YHO in the monthly bulletins from January till May of 2004. Oyak Pension fund codes were matched with relative ING Pension fund codes until December 2008. Koç Allianz fund codes were matched with relative Allianz fund codes for the same time period. Doğan Pension fund codes were matched with relative Fortis Pension fund codes until October 2005. Commercial Union fund codes were matched with relative Aviva fund codes until August 2004. Dışbank fund codes were matched with relative Fortis Bank fund codes until November 2005. C Bank fund codes were matched with relative Bankpozitif fund codes until January 2006. Bankeuropa fund codes were matched with relative Millennium Bank fund codes until October 2007. And finally Oyakbank fund codes were matched with relative ING Bank fund codes until May 2008.

8.3 Appendix C - Survivorship Bias

To check whether survivorship bias has an important role in Turkish markets for the period of January 2004 to December 2009, we first use Blake and Timmermann (1998)'s measure. Here, $r_{s,j,t}$ represents equally weighted portfolio return of all the funds that survive at the observation period without looking at the issue date, for month t, at eq. (12). The results indicate a 0.68% per annum survivorship bias which is negligible compared to UK as discussed earlier.

Table 14: Gruber's survivorship bias measure results

Method: Least Squares
Sample: 2004M02 2009M10
Included observations: 69

	Variable	Coefficient	Std. Error	t-Statistic	Prob.
All Funds	ISE100-RF	0.519950	0.027989	18.57723	0.0000
	C	-0.002400	0.001528	-1.570525	0.1210
Non-surviving Funds	ISE100-RF	0.536607	0.034603	15.50734	0.0000
	C	-0.002066	0.003577	-0.577459	0.5656

Second measure discussed is of Gruber's. Gruber (1996) in his study examined the regression results of average excess returns of all funds over the excess market returns and average excess returns of non-surviving funds over the excess market returns. Implementing this procedure, we first regressed non-surviving funds' average excess returns on the excess market returns for our sample period (excluding the last two months since there are no non-surviving funds.) Then we regressed all funds' average excess returns on the excess market returns for the same sample period and compared the results. From table (14) it is seen that both alphas of the regressions are around -0.2% monthly and are insignificant. The difference between these alphas annually is 0.36% which is negligible. Hence we can say that there is no survivorship bias for our sample period.

Finally, we regressed 139 surviving funds' average excess returns on the 59 non-surviving funds' average excess returns with the null hypothesis that alpha, the extra performance is zero. Our alpha is 0.000378 with probability 0.87 which means that we cannot reject alpha being zero. We conclude that surviving funds do not generate extra performance over non-surviving funds. Hence we can assume that there is no survivorship bias in Turkish markets for our sample period.

Table 15: Survivorhip bias for Turkish market

Dependent Variable: ALL_SURVIVING-RF
Method: Least Squares
Sample (adjusted): 2004M02 2009M10
Included observations: 69 after adjustments
Newey-West HAC Standard Errors & Covariance (lag truncation=3)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
NON_SURVIVING-RF	0.801892	0.113426	7.069728	0.0000
C	0.000378	0.002384	0.158415	0.8746

R-squared	0.837777	Mean dependent var	0.001373
Adjusted R-squared	0.835356	S.D. dependent var	0.051624
S.E. of regression	0.020947	Akaike info criterion	-4.865059
Sum squared resid	0.029399	Schwarz criterion	-4.800302
Log likelihood	169.8445	F-statistic	346.0119
Durbin-Watson stat	1.043619	Prob(F-statistic)	0.000000

9. Reference

Abdel-Kader, M. & Qing, K. Y. 2007, 'Risk-adjusted performance, selectivity, timing ability and performance persistence of Hong Kong mutual funds', *Journal of Asia-Pacific Business*, Vol. 8, No. 2, pp. 25-58.

Blake, D. & Timmermann, A. 1998, 'Mutual fund performance: evidence from the UK', *European Finance Review*, Vol. 2, No. 1, pp. 57-77.

Cai, J., Chan, K. C. & Yamada, T. 1997, 'The performance of Japanese mutual funds', *The Review of Financial Studies*, Vol. 10, No. 2, pp. 237-273.

Capital Markets Board of Turkey, 1996, *Yatırım Fonlarına İlişkin Esaslar Tebliği*, Retrieved from: http://www.spk.gov.tr/mevzuat/pdf/Bolum_4/4_7_1_1_SeriVII_No10.pdf (02.08.2010)

Capital Markets Board of Turkey, 2001, *Bireysel Emeklilik Tasarruf ve Yatırım Sistemi Kanunu*, Retrieved from: http://www.spk.gov.tr/displayfile.aspx?action=displayfile&pageid=288&fn=288.pdf (02.08.2010)

Capital Markets Board of Turkey, 2002, *Emeklilik Yatırım Fonlarının Kuruluş ve Faaliyetlerine İlişkin Esaslar Hakkında Yönetmelik*, Retrieved from: http://www.spk.gov.tr/mevzuat/pdf/Bolum_4/4_7_1_EYF_Yon.pdf (02.08.2010)

Capital Markets Board of Turkey, 2008, *Sermaye Piyasası Kanunu*, Retrieved from: http://www.spk.gov.tr/displayfile.aspx?action=displayfile&pid=0&menuid=5&subid=0 (02.08.2010)

Carhart, M. M. 1997, 'On persistence in mutual fund performance', *Journal of Finance*, Vol. 52, No. 1, pp. 57-82.

Chen, L., Jegadeesh, N. & Wermers, R. 2000, 'The value of active mutual fund management: an examination of the stock holdings and trades of fund managers', *Journal of Financial & Quantitative Analysis*, Vol. 35, No. 3, pp. 343-368.

Christensen, M. 2005, 'Danish mutual fund performance selectivity, market timing and persistence', *Finance Research Group Working Papers*, F-2005-01, University of Aarhus, Aarhus School of Business, Department of Business Studies.

Cuthbertson, K., Nitzsche, D. & O'Sullivan, M. 2004, 'UK mutual fund performance: genuine stock picking ability or luck', *Money Macro and Finance (MMF) Research Group Conference 2004*, Retrieved from: http://repec.org/mmfc04/55.pdf (11.20.2009)

Dahlquist, M., Engström, S. & Söderlind, P. 2000, 'Performance and characteristics of Swedish mutual funds', *The Journal of Financial and Quantitative Analysis*, Vol. 35, No. 3, pp. 409-423.

Daniel, K., Grinblatt, M., Titman, S. & Wermers, R. 1997, 'Measuring mutual fund performance with characteristic-based benchmarks', *Journal of Finance*, Vol. 52, No. 3, pp. 1035-1058.

Daniel, K., Grinblatt, M., Titman, S. & Wermers, R. 1997, 'Measuring mutual fund performance with characteristic-based benchmarks', *Journal of Finance*, Vol. 52, No. 3, pp. 1035-1058.

EFAMA, 2009, *International quarterly statistics q4 2009 supplementary tables*, Retrieved from: http://www.efama.org/index2.php?option=com_docman&task=doc_view&gid=1173&Itemid=-99 (03.03.2010)

EFAMA, 2009, *International quarterly statistics*, Retrieved from: http://www.efama.org/index2.php?option=com_docman&task=doc_view&gid=1174&Itemid=-99 (03.03.2010)

Fama, E. F. & French, K. R. 2008, 'Mutual fund performance', *National Bureau of Economic Research (NBER)*, working paper, Retrieved from: http://ssrn.com/abstract=1153715 (09.10.2009)

Fama, E. F. & French, K. R. 2009, 'Luck versus skill in the cross section of mutual fund returns', *Tuck School of Business Working Paper, No. 2009-56, Chicago Booth School of Business Research Paper*, Retrieved from: http://ssrn.com/abstract=1356021 (09.10.2009)

Fama, E.F. & French, K. R. 1992a, 'The cross-section of expected stock returns', *Journal of Finance*, Vol. 47, No. 2, pp. 427-465.

Fama, E.F. & French, K. R. 1993, 'Common risk factors in the returns on stocks and bonds', *Journal of Financial Economics*, Vol. 33, No. 1, pp. 3-56.

Fama, E.F. & French, K. R. 1996, 'Multifactor explanations of asset pricing anomalies', *Journal of Finance*, Vol. 51, No. 1, pp. 55-84.

Ferson, W. E. & Schadt, R.W. 1996, 'Measuring fund strategy and performance in changing economic conditions', *Journal of Finance*, Vol. 51, No. 2, pp. 425-461.

Grant, D. 1977, 'Portfolio performance and the "cost" of timing decisions', *Journal of Finance*, Vol. 32, No. 3, pp. 837-846.

Grinblatt, M. & Titman, S. 1989, 'Mutual fund performance: an analysis of quarterly portfolio holdings', *Journal of Business*, Vol. 62, No. 3, pp. 393-416.

Hendricks, D., Patel, J. & Zeckhauser, R. 1993, 'Hot hands in mutual funds: short run persistence of relative performance, 1974-88', *Journal of Finance*, Vol. 48, No. 1, pp. 93-130.

Henriksson, R. D. & Merton, R. C. 1981, `On market timing and investment performance. II. statistical procedures for evaluating forecasting skills`, *Journal of Business*, Vol. 54, No. 4, pp. 513-533.

Jegadeesh, N. & Titman, S. 1993, 'Returns to buying winners and selling losers: implications for stock market efficiency', *Journal of Finance*, Vol. 48, No. 1, pp. 65-91.

Jensen, M. C. 1967, 'The performance of mutual funds in the period 1945-1964', *Journal of Finance*, Vol. 23, No. 2, pp. 389-416.

Kosowski, R., Timmermann, A., Wermers, R. & White, H. 2005, 'Can mutual fund stars really pick stocks? new evidence from a bootstrap analysis', working paper, Retrieved from: http://ssrn.com/abstract=855425 (11.17.2009)

Mazuy, K. K. & Treynor, J. L. 1966, 'Can mutual funds outguess the market?', *Harvard Business Review*, Vol. 44, No. 4, pp. 131-136.

Sharpe, W. F. 1966, `Mutual fund performance`, *Journal of Business*, Vol. 39, No. 1, pp. 119-138.

Treynor, J. L. 1965, `How to rate management of investment funds`, *Harvard Business Review*, Vol. 43, No. 1, pp. 63-75.

Efron, B. 1998, `R. A. Fisher in the 21st century`, *Statistical Science*, Vol. 13, No. 2, pp. 95-122.

Efron, B. 1979, `Bootstrap methods: another look at the jackknife`, *The Annals of Statistics*, Vol. 7, No. 1, pp. 1-26.

Efron, B. & Gong, G. 1983, `A leisurely look at the bootstrap, the jackknife and cross-validation`, *The American Statistician*, Vol. 37, No. 1, pp. 36-48.

Quenouille, M. H. 1949, `Approximate tests of correlation in time-series`, *Journal of the Royal Statistical Society, Series B (Methodological)*, Vol. 11, No. 1, pp. 68-84.

Efron, B. & Tibshirani, R. J. 1993, *An introduction to the bootstrap*, Chapman and Hall, New York.

Efron, B. 1982, *The jackknife, the bootstrap, and other resampling plans*, Capital City Press, Montpelier, Vermont, USA.

Efron, B. & Tibshirani, R. 1986, `Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy`, *Statistical Science*, Vol. 1, No. 1, pp. 54-77.

Chernick, M. R. 1999, *Bootstrap methods a practitioner's guide*, John Wiley & Sons, Canada.

Davidson, R. & MacKinnon, J. G. 2004, *Econometric theory and methods*, Oxford University Press, New York.

Mooney C. Z. & Duval, R. D. 1993, *Bootstrapping: A nonparametric approach to statistical inference*, Sage University Paper series on Quantitative Applications in the Social Sciences, 07-095, Newbury Park, CA.

Miller, R. G. 1974, `The jackknife--a review`, *Biometrica*, Vol. 61, No. 1, pp. 1-15.

Efron, B. & Tibshirani, R. 1995, `Cross-validation and the bootstrap estimating the error rate of a prediction rule`, Technical report 176,

Department of Statistics, Stanford University, retrieved from:
http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.35.9840
(11.19.2009)

Bickley, P. J. & Freedman, D. A. 1981, `Some asymptotic theory for the bootstrap`, *The Annals of Statistics*, Vol. 9, No. 6, pp. 1196-1217.

Singh, K. 1981, `On the asymptotic accuracy of efron's bootstrap`, *The Annals of Statistics*, Vol. 9, No. 6, 1187-1195.

Hinkley, D. V. 1988, `Bootstrap methods`, *Journal of the Royal Statistical Society*, Vol. 50, No. 3, pp. 321-337.

Athreya, K. B. 1987, `Bootstrap estimation of the mean in the infinite variance case`, *The Annals of Statistics*, Vol. 15, No. 2, pp. 724-731.

Gine, E. & Zinn, J. 1989, `Necessary conditions for bootstrap of the mean`, *The Annals of Statistics*, Vol. 17, No. 2, pp. 684-691.

Silverman, B. W. & Young, G. A. 1987, `The bootstrap: to smooth or not to smooth?`, *Biometrika*, Vol. 74, No. 3, pp. 469-479.

Carlstein, E. 1986, `The use of subseries values for estimating the variance of a general statistic from a stationary sequence`, *The Annals of Statistics*, Vol. 14, No. 3, pp. 1171-1179.

Künsch, H. R. 1989, `The jackknife and the bootstrap for general stationary observations`, *The Annals of Statistics*, Vol. 17, No. 3, pp. 1217-1241.

Freedman, D. A. 1981, `Bootstrapping regression models`, *The Annals of Statistics*, Vol. 9, No. 6, pp. 1218-1228.

10. Notes

[1] Turkish Capital Markets Law, Article 37.

[2] Turkish Capital Markets Law, Article 38.

[3] Communiqué on Principals Regarding Mutual Funds, Article 42.

[4] Communiqué on Principals Regarding Mutual Funds, Article 5.

[5] Since we don't have any non-surviving funds for the last two months, for comparability those months are excluded and we have 69 observations instead of 71.

[6] Preferred stocks have higher claims to the firm's assets and earnings relative to common stocks. They usually don't have voting rights and they have priorities on dividend payments.

[7] Relevant Matlab codes for bootstrap simulation are available upon request.

[8] www.fonbul.com