THE INVESTMENT PERFORMANCE OF A PENSION FUND′S INDIVIDUAL INVESTORS: A BEHAVIORAL FINANCE APPROACH

YİĞİT OKAR

BOĞAZİÇİ UNIVERSITY

2018

THE INVESTMENT PERFORMANCE OF A PENSION FUND′S INDIVIDUAL INVESTORS: A BEHAVIORAL FINANCE APPROACH

Thesis submitted to the

Institute for Graduate Studies in Social Sciences

in partial fulfillment of the requirements for the degree of

Master of Arts

in

Economics

by

Yiğit Okar

Boğaziçi University

2018

DECLARATION OF ORIGINALITY

I, Yiğit Okar, certify that

- I am the sole author of this thesis and that I have fully acknowledged and documented in my thesis all sources of ideas and words, including digital resources, which have been produced or published by another person or institution;
- this thesis contains no material that has been submitted or accepted for a degree or diploma in any other educational institution;
- this is a true copy of the thesis approved by my advisor and thesis committee at Boğaziçi University, including final revisions required by them.

Signature......................................

Date....28.06.2018....................

ABSTRACT

The Investment Performance of a Pension Fund's Individual Investors:

A Behavioral Finance Approach

In this thesis, we investigated the portfolio choice and trading characteristics of individuals using data from pension accounts. The data contains nearly 17000 individuals' accounts during the years 2008-2013. Various and seemingly inefficient patterns are observed. We observed that as the time flows, average equity allocation increases. Not-married people significantly hold more equity in their portfolios. Individuals who are middle aged(35-44) hold equity in their portfolio most. We could not find a difference in the equity allocation characteristics between different genders. As for trading habits, males and married people significantly trade more than their opposites. In addition, people who are between age 44-54 trade significantly more. Moreover we observed high serial correlation in portfolio change decisions. Briefly, equity allocation and trading decisions are affected by demographic characteristics, also there is serial correlation in that decisions. These habits constitute *behavioral biases*. Furthermore, we assess the performances obtained by the individual pension fund investors. Majority of the people in the plan do not take any risk, thus obtain low return but high Sharpe ratios.

We also studied return predictability in an aim to improve the performance of the portfolios by employing a simple trading algorithm. In general, we found little

predictability in short term using classical methods. We observed that as we increase the investment horizon, the predictability increases. We claim that individual investors who lack the financial intellectuality and have aforementioned behavioral biases, could use the help of automated algorithms to ameliorate the performance of their portfolios.

ÖZET

Emeklilik Hesabı Yatırımcılarının Performans Analizi:

Bir Davranışsal Finans Yaklaşımı

Bu tezde, emeklilik hesaplarından elde edilen verileri kullanarak kişilerin portföy seçimini ve alım satım özelliklerini inceledik. Veriler, 2008-2013 yılları arasında yaklaşık 17000 bireyin hesabını içermektedir. Çeşitli ve verimsiz gözüken yaklaşımlar gözlenmiştir. Bireylerin çoğu varlıklarının herhangi bir bölümünü hisse senetlerine ayırmamıştır. Zaman ilerledikçe ortalama hisse senedi tahsisinin arttığı gözlemlenmiştir. Evli olmayan insanlar portföylerinde daha fazla hisse senedi fonu tutarlar. Literatürle paralel olarak, orta yaşlı bireyler (35-44) en çok portföyünde hisse tutan gruptur. Farklı cinsiyetler arasındaki hisse tutuş miktarlarında bir farklılık gözlemlenmemektedir. Al-sat alışkanlıklarına gelince, erkekler ve evli insanlar, karşıtlarından çok daha fazla al-sat yapıyorlar. Buna ek olarak, 44-54 yaş arasındaki insanlar önemli ölçüde daha fazla al-sat yapmaktadirlar. Ayrıca, emeklilik hesaplarının yönetiminde bir başka verimsizlik kaynağı olan portföy değişikliği kararlarında yüksek düzeyde bir oto-korelasyon gözlemledik. Buna ek olarak bireysel emeklilik fonu yatırımcılarının performanslarını değerlendirdik. Plandaki insanlarn çoğunluğu herhangi bir risk almamakta, düşük getiri elde etmekte ancak yüksek Sharpe oranları elde etmektedir.

Ayrıca, davranışsal verimsizliklerden kurtulabilmeye ve portföylerin performansını iyileştirmeye imkan sağlayacak basit bir al-sat algoritması kurgulamak amacıyla getiri öngörülebilirliği üzerinde çalıştık. Genel olarak, klasik yöntemlerle kısa vadede çok az öngörülebilirlik bulduk. Yatırım ufkunu arttırdıkça, öngörülebilirlik literatürüne paralel olarak öngörülebilirliğin arttığını gözlemledik. Finansal entelektüelliği olmayan ve davranışsal olarak verimsiz olan bireysel yatırımcıların, portföylerinin performansını iyileştirmek için otomatik algoritmaların yardımının kullanılabileceğini öneriyoruz.

ACKNOWLEDGEMENTS

Firstly I would like to thank my thesis advisor Burak Saltoğlu for leading me, sharing his ideas and giving me the opportunity to work with him. Additionally, I would like to thank Economics Department of Boğaziçi University for being there whenever I needed them to be.

I would also like to thank Ayhan Yuksel for providing me his ideas on the subject and his supportive manner. He not only thought me a lot but also motivated me greatly. Moreover I would like to thank Tolga Umut Kuzubaş for being a great supervisor. His door to his office was always open to me and he explained the material in great detail to me whenever I needed.

I express my special thanks to Günnur Ege Bilgin for providing me her help. I would like to thank the Scientific and Technological Research Council of Turkey (TÜBİTAK) BİDEB for the financial support through the 2210-A Master's Scholarship Program.

Of course, I am also deeply thankful to my family. This accomplishment would not have been possible without them. Thank you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The performance of individual investors are studied widely in the literature. In this study, we try to analyze the main decisions which are equity allocation decisions and trading decisions, taken by a pension fund's investors. We investigate these behaviours and link them to investors' demographic characteristics. Additionally we try to detect other patterns in the decisions of investors including market timing, trend following or auto-correlation in these decisions. We then look at some performance metrics that individual investors realize in the pension plan. Moreover, we study return predictability using data from Turkey and aim to improve the performance of the individuals' by employing a simple trading strategy, thus omitting various behavioral biases (equity allocation and trading decisions are affected by demographic characteristics, serial correlation in these decisions, people do not take any risk by holding equities.).

In Odean (1999) it is claimed that some fraction of the people trade too much which reduces their overall return. Furthermore, he proposes that the ones who trade more should be the overconfident ones and people who have brokerage accounts fall in this category. He concludes that even the trading costs are excluded from the study the overconfident investors lower their returns through trading. A similar conclusion is made in Barber and Odean (2000). In that paper they show that the investors who trade most earn an annual return of 11% where the market return is 17%. Note that in both of the studies, the data includes individuals who own brokerage accounts. In Agnew et. al(2003), they pursue a similar study using a large 401(k) plan which is very close to the data used in this thesis. They find some patterns in equity allocation and trade decisions, similar to Odean(1999) and Barber and Odean(2000). In this thesis, we aim to find similar patterns in trading and equity allocation decisions which may be causing inefficiencies in terms of fund performance. Moreover, we test -with a simple trading algorithm- that whether these inefficiencies can be removed.

Implementing a trading strategy for equities is tricky using data from Turkey because Turkey may fail to have Equity Risk Premium despite the current perception in the literature about Equity Risk Premiums in Emerging Markets(Donadelli and Prosperi (2012), Salomons and Grootveld (2002)).

The literature on the return predictability is mixed. Fama and French (1989) assert that there is mounting evidence that stock returns are predictable. They use variables like Dividend Yield to predict returns and relate the predictability to the fact that these variables reflect the long-term business cycles which obviously affect the returns. Campbell and Shiller (1988) use a VAR framework and variables like earning-price ratios and dividend-price ratios in order to predict returns. Ferson and Harvey (1993) use a unique approach to return prediction problem as they focus on the marginal effect of local market variables while controlling for global variables. They also relate the source of predictability to the global risk premia. Cochrane(1999) also assert that returns are predictable via using variables like dividend/price ratio especially over the long horizons. In addition he claims that shorter term- like daily or weekly-returns are close to unpredictable. Timmerman and Cenezsizoglu (2012) handle the return predictability in a different manner. They try to forecast the returns using both mean and variance estimates rather than just a mean estimate. Moreover, they claim that while forecasting using return predictability models can underperform historical mean model, making investment decisions by looking at these forecasts add economic value.

On the contrary, in their comprehensive study Goyal and Welch (2008) claim that most of the methods performs poorly when they try to make out of sample forecasts or use the predictive models with more data. They conclude that most of the models are unstable and spurious.

CHAPTER 2

PORTFOLIO CHOICE AND TRADING

In this chapter we investigate whether there is investor biases that jeopardize the fund performance or not by looking at their trading activity, equity allocation and portfolio performance. Moreover we study characteristics of individual investors' fund performance using data from a pension fund. In Odean (1999), it is stated that people who have brokerage accounts turn out to be over confident and this lead them to excess trading activity which decreases investor returns. Additionally, it is claimed that men significantly trade more than women. However, these studies examine a narrow sample of the population. In a more comprehensive study, Agnew et al. (2003), a larger sub-sample is investigated. A 401(k) plan consists of 7,000 accounts through 1994-1998. Several interesting conclusions are made in their paper. First, they find that the equity allocations are at extremes. That is individuals either hold too much equity or zero equity. Additionally, they concluded that married and high income investors tend to have more stocks in their portfolios. Finally, they found out that trading is infrequent in the 401(k) plan they investigate.

We pursue a study to gauge if these concepts are present in the Turkey as well. We use data from pension accounts. We first examine the descriptive statistics of the participants to better understand how demographics affect the allocation and trading behaviour. Then we run regressions to see how these factors jointly affect the decisions taken. Moreover we look at the market timing, trend following abilities and serial correlation of decisions taken of individuals in the data. Finally, we investigate the returns of funds obtained by individual investors in a 6 year period of 2008-2013.

This chapter is organized as follows. In section 2.1 the data used is summarized. In 2.2, equity allocations of individuals are investigated. In section 2.3 trading activity and the characteristic that affect this behavior is questioned. Moreover, in that section, we test for either participants try to do market timing or trend following. In section 2.4 we study the individuals' return performance.

2.1 Data

The data in this thesis come from individual pension fund accounts. It is called Individual Retirement System (IRS) which started in 2003. The data contains approximately 180,000 individual accounts. However we excluded the participants who was not in the plan in the years between 2008-2013. After the elimination we end up with 17,298 participants who stay in the plan for 6 years. The data includes detailed information on participants demographics. Gender, age, marital status, education, earnings are included in the data set.

The demographic characteristics and descriptive statistics of the individuals are presented in Table 1. Age is reported as of 2008. The data is divided by 57.8% male and 42.2% female. The majority of the participants are married (77.20 percent). Average income is 47850 with a standard deviation of 33992. The average age of the individuals in the sample is 38.7. Approximately half of the participants have at least undergraduate degree and the other half has primary or high school education.

Table 1. Descriptive Statistics for Those Who Stay in the Plan for 6 Years

	Observations	Percent	Mean	Standard Deviation	Minimum	Maximum
Gender:						
Male	9985	57.8 %				
Female	7313	42.2 %				
Total	17298	1				
Marital Status:						
Married	13363	77.20%				
Single	3934	22.80 %				
Unknown	1	≈ 0.00 %				
Total	17298	1				
Income (Yearly):	17298		47850	33992	6996	399418
Age:	17298		38.07	8.6	19	77
Education:						
Primary	2820	16.30%				
High School	5412	31.28 %				
College	7601	43.94 %				
Post College	1465	8.46%				
Total	17298	1				

2.2 Equity allocations

For the purpose of this study we divide assets in to two goups, namely, risky and non-risky assets. Risky assets consists of equities and non-risky assets consists of other instruments. Correspondingly we classify individuals' risk appetite according to how much equity they hold. Table 2 shows the yearly equity allocation statistics.In every grouping in Panel B the fraction of the data that is marked by bold typing is considered a base year. Using this base year we conduct a Mann-Whitney mean equivalence test in order to asses the significance. In the first panel the distribution is summarized. One immediate observation is that 62% of the observations does not include any equity in their portfolios. And as the percentage of equity increases, the fraction of people holding that much of equity decreases.

Equity allocations vary over time. From 2008 to 2009 we see a slight decrease in average equity holded in portfolios. After 2009 we observe a significant increase in the equity allocation. The participants may be responding to sub-prime crisis by first decreasing their average equity holdings and than increasing to benefit the post crisis increase.

Not married people significantly hold more equity, meaning that they take more risk. This fact is on the contrasting to the literature. In Agnew et al. (2003) it is stated that married people take more risk by holding more equities in their portfolio. Agnew et al. state the reason of such a behavior is the fact that married people has two sources of income thus they have an opportunity to take risk. In our study, the situation is the opposite. There can be many explanations to this phenomenon. One of them could be that the married couples may be thinking about their born or unborn children thus does not want to take risk.

Next, we question how asset allocation changes with age. An inverse U shaped (or hump shaped) curve for the age is observed. Equity allocation, in the beginning, is increasing as a function of age, then declines. People under 35 holds 5.41 % equity in their portfolios wheras age group 35-44 holds 6.11 %. From here it starts to decline as age groups 45-54 hold 5.82% , 55-64 5.51% and 65+ hold 5.50%. This concept is parallel with the models of optimal portfolio choice. As investors age they tend to take less risk since they do not have the same future income stream ahead of them. This phenomennon is shown by Jagannathan and Kocherlachota (1996). Note that there are various papers that speculate on the reason of this concept.

Afterwards, we observe that there is a positive relationship between the income and the equity holdings. As investors' labor income increase they have more space to take risk thus they increase their average equity holdings. Salary group $\leq$ TRY 25000 holds an average of 4.90 % equity. From that point average equity holdings increase monotonically to 8.49% associated with the salary group $\geq$ TRY 100000.

Note that we could not find any evidence on the fact that equity holdings vary with gender.

Table 2. Equity Allocations * –> 5 Percent ** –> 1 Percent

Panel A: Distribution

Range	Percent
x = 0	62.6 %
0 < x < 10	14.8 %
10 < x < 20	12.7 %
20 < x < 40	8.20 %
40 < x < 80	1.50 %
80 < x < 100	0.14 %

Panel B: Statistics by Group

	Observation	Mean	Standard Deviation
All	103788	5.78 %	10.48%
Sort by year:			
2008	17298	4.97%	8.86%
2009	17298	4.29%**	8.36%
2010	17298	4.84%	8.97%
2011	17298	6.77%**	11.69%
2012	17298	6.93%**	12.22%
2013	17298	6.86%**	11.73%
Gender:			
Male	59910	5.95%	10.95%
Female	43878	5.53%	9.80 %
Marital Status:			
Married	80178	5.68%	10.39%
Not Married	23610	6.11%**	10.79 %
Age:			
Under 35	29361	5.41 %	10.23%
35-44	41040	6.06%**	10.83%
45-54	26252	5.82%**	10.42%
55-64	6529	5.51%*	9.74%
65+	606	5.50%	8.79%
Annual Salary:			
<**25000**	30066	4.90 %	9.69%
25000-49999	36222	5.58%**	10.39%
50000-74999	19164	6.00%**	10.53%
75000-99999	10434	6.50%**	10.65%
100000+	7902	8.49%**	12.65%

We have used the censored regression model for panel data with individual specific effects which has the following specification. The regression results are presented in Table 3. As dependent variables, we have used an is-married dummy which is reported as 'married'. Similarly we use a gender dummy variable 'gender', symbolizes male participants when equal to one.We also employed age, income and year dummies.

$$y_{it}^{*} = x_{it}^{'}\beta + \epsilon = x_{it}^{'}\beta + \mu_i + \nu_{it}$$

$$y_{it} = \begin{cases} 0 & \text{if } y_{it}^{*} \leq 0 \\ y_{it}^{*} & \text{if } 0 < y_{it}^{*} < 1 \\ 1 & \text{if } y_{it}^{*} \geq 1 \end{cases}$$

Where y_{it} is the amount of equity allocation for individual i at time t. One immediate fact is that the estimated coefficients are very small because our dependent variable generally (more that 50 % of the data) takes values between $0.00 - 0.10$. Firstly, note that we see a positive trend in equity allocations by time via looking the the dummy variables. Secondly, parallel with our findings in Table 2, being married decreases the amount of average equity holdings. Note that when we control for other demographic characteristics the effect of income seems to disappear. As, expected the coefficient of the variable 'age' is insignificant.

Table 3. The Censored Regression Estimation for Panel Data

	Model
(Intercept)	-0.091375^{***}
	(0.002791)
married	-0.009642^{***}
	(0.001234)
gender	-0.012515^{***}
	(0.001051)
age	0.000119
	(0.000062)
income	0.000000^{***}
	(0.000000)
2009	-0.018303^{***}
	(0.003125)
2010	-0.007602^{***}
	(0.002307)
2011	0.041229^{***}
	(0.002146)
2012	0.043210^{***}
	(0.001972)
2013	0.049419^{***}
	(0.001783)
logSigmaMu	-1.797953^{***}
	(0.003206)
logSigmaNu	-2.299783^{***}
	(0.001401)
AIC	-5145.047425
BIC	-5030.446157
Log Likelihood	2584.523712
Num. obs.	103788
Left-censored	65020
Uncensored	38768
Right-censored	0

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

2.3 Number of trades

A similar study that is similar to that we do about equity allocations is made about annual trading activity. We defined that when an individual changes the allocation of her portfolio or add/remove a fund from a portfolio, she is said to be traded. We again first look at the general distribution of annual trades made by participants of the plan, then we characterize these behaviours by demographics. Finally, we use logistic regression in order to supplement the study. Table 4 presents our findings.

Similar to previous chapter, panel A describes the distribution. An immediate observation is that $\approx 65\%$ of the observations make 0 annual yearly trade. This finding is the opposite to the conclusion made by Odean (1999). We speculate that the reason behind this difference could be the data sources. They get their data from a brokerage account while we take ours from pension accounts of individuals. It is perfectly sensible that people tend to not trade their pension accounts aggressively as they may not want to take risk on their pension accounts. A similar conclusion is made in Agnew et al. (2003). They connect the cause of limited trading activity to limited fund choices in 401(k) plans, which could be valid in our case as well.

In panel B we investigate the patterns in number of trades made by participants. There is a positive time trend in trading activity with one exeption, 2012. Males seem to trade significantly more frequently than females. Males trade 2.12 times on average while females trade 1.92 times annually. Being married is also significantly effective on the number of trades. While married people trade 2.07 times on average, not-married people tend to trade 1.93 times per year. Trading activity monotonically and significantly increases with age. People who are under 35 tend to trade 1.46 times annually and number of trades per year increases to 3.02 times for the people who are ≥ 65. As stated in Agnew et al. (2003), this phenomenon can be explained by the fact that as people age their present value of human capital earnings decrease. Additionally, their financial wealth increases. Thus rebalancing has a more substantial effect on the older people.

Finally, as the annual earnings of the individuals increase, the number of trade increases. People who earn less than TRY 25000 trade their accounts 1.86 times on average. From that point, trading activity increases to 2.43 times on average for people who earn greater than TRY 100000. Again in Agnew et al. (2003), a similar conclusion is made. People with more earnings have a larger financial wealth for whom the importance is rebalancing is greater.

Table 4. Number of Trades * –> 10 Percent ** –> 5 Percent *** –> 1 Percent

Panel A: Distribution

Range	Percent
x = 0	63.52 %
x = 1	6.16 %
$2 \leq x \leq 5$	10.03 %
$5 < x \leq 10$	18.72 %
$10 < x$	1.55 %

Panel B: Statistics by Group

	Observation	Mean	Standard Deviation
All	103788	2.041	3.34
Sort by year:			
2008	17298	0.7196208	1.979577
2009	17298	0.79275068*	2.111343
2010	17298	0.8026361	2.195564
2011	17298	1.1911204***	2.198668
2012	17298	0.6880564	1.998305
2013	17298	8.0508729***	1.192934
Gender:			
Male	59910	2.127641	3.379152
Female	43878	1.922330**	3.283455
Marital Status:			
Married	80178	2.070605	3.358417
Not Married	23610	1.939771**	3.277218
Age:			
Under 35	29361	1.465005	2.913690
35-44	41040	2.164254**	3.418693
45-54	26252	2.320433**	3.481239
55-64	6529	2.639302**	3.655960
65+	606	3.023102**	3.763170
Annual Salary:			
<**25000**	30066	1.860407**	3.246875
25000-49999	36222	2.000497**	3.318418
50000-74999	19164	2.147255**	3.400784
75000-99999	10434	2.207207**	3.434731
100000+	7902	2.434574**	3.465703

We run a logistic regression in order to asses the importance of demographics in the trade decisions individuals face in each month. Since we want to determine that if the demographics like age, gender, being married or income effect the probability of making a trade in a specific month, we used logistic regression. The model is as follows:

$$Trade = \sigma(X\beta) + \epsilon$$

where σ represents the sigmoid function. The independent variables used in the regression are $gender$, age, $married$, $income$ and $year\ dummies$.

Table 5 presents the results of the logistic regression, Table 6 presents the marginal effects and Table 7 presents the confusion matrix associated with the model. First of all, when we look at the confusion matrix we see a relatively good performance with an accuracy of 0.87. The independent variables are all significant and shows a positive time trend until 2013 which verify the conclusions made in the descriptive statistics. Being male has a significantly positive effect on the trading activity. Again parallel with our findings in the Table 4, being married also has a positive effect on the trading activity. Interestingly, when we control for demographics the effect of $income$ on trading activity seems to disappear, similar to our findings for the equity allocations. According to regression results, Age effects the trading activity positively which again is confirming the conclusions made in the Table 4. We conclude that in our sample which includes pension accounts of $\approx 17,000$ individuals, earning high income does not affect the equity allocation decisions and trading activity after controlling for other demographic characteristics.

Table 5. Logistic Regression Results

	Dependent variable:
	monthly_trade
married	0.083^{***}
	(0.008)
gender	0.096^{***}
	(0.006)
age	0.002^{***}
	(0.0004)
income	0.00000^{***}
	(0.00000)
2009	0.102^{***}
	(0.013)
2010	0.114^{***}
	(0.013)
2011	0.543^{***}
	(0.012)
2012	-0.054^{***}
	(0.013)
2013	3.474^{***}
	(0.011)
Constant	-3.130^{***}
	(0.017)
Observations	1,245,456
Log Likelihood	-391,206.500
Akaike Inf. Crit.	782,433.100
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 6. Marginal Effects

	factor	AME	SE	z	p	lower	upper
1	age	0.00014	0.00003	4.26080	0.00002	0.00008	0.00020
2	income	0.00000	0.00000	44.47822	0.00000	0.00000	0.00000
3	gender	0.00866	0.00058	14.91962	0.00000	0.00752	0.00979
4	2009	0.00922	0.00116	7.95976	0.00000	0.00695	0.01149
5	2010	0.01029	0.00116	8.89079	0.00000	0.00802	0.01255
6	2011	0.04915	0.00108	45.69617	0.00000	0.04704	0.05126
7	2012	-0.00488	0.00120	-4.05347	0.00005	-0.00724	-0.00252
8	2013	0.31428	0.00084	372.14911	0.00000	0.31262	0.31593
9	married	0.00748	0.00069	10.90245	0.00000	0.00614	0.00883

Table 7. Confusion Matrix of Fit

	Predicted no trade	Predicted Trade
No trade	0.78	0.05
Trade	0.06	0.11

Additionally, we tested if the participants try to time the market or they act according to the market known as feedback trading. Table 8 presents the the correlation coefficients estimated. When we look at the correlation between one-two-three month lagged and lead return of BIST100 and monthly change of equity allocations we can not see a significant correlation.

However we observed significant auto-correlation in changes of equity proportions. The correlation of one month lagged changes in equity proportion is 0.23 which displays positive serial correlation. This notion is consistent with the notion that individual investors are changing their portfolios inefficiently. For example, they pay unnecessary transaction fees. Additionally, they are unnecessarily hesitant, which jeopardize the future unrealized returns.

Table 8. Correlation Coefficients

Correlation	Coefficient
$\rho(\Delta_{equity\ allocation}, r_{t-1,t})$	-0.0088
$\rho(\Delta_{equity\ allocation}, r_{t-2,t})$	-0.0043
$\rho(\Delta_{equity\ allocation}, r_{t-3,t})$	0.0072
$\rho(\Delta_{equity\ allocation}, r_{t+1,t})$	0.0197
$\rho(\Delta_{equity\ allocation}, r_{t+2,t})$	0.0160
$\rho(\Delta_{equity\ allocation}, r_{t+3,t})$	-0.0260
$\rho(\Delta^{t}_{equity\ allocation}, \Delta^{t-1}_{equity\ allocation})$	0.2348
$\rho(\Delta^{t}_{equity\ allocation}, \Delta^{t-2}_{equity\ allocation})$	0.2133
$\rho(\Delta^{t}_{equity\ allocation}, \Delta^{t-3}_{equity\ allocation})$	0.3720

2.4 Individual portfolio performances

In this section we look at the returns obtained by individual investors. Table 9. presents the distribution of return performance metrics. As expected the majority of people have low returns and low standard deviaton in returns since the majority of the people in the plan do not hold equity. 70 % of the individuals hold equities no more than 1 % and 80 % of the individuals hold equities no more than 5 % as described in section 2.2. Thus these individuals have very low standard deviation in returns which lead to high Sharpe ratios. We evaluated 17298 individuals' return performance by looking to their total return (Overall Return in the Table 9) in 6 years, Sharpe ratios, compound annual growth rates, mean of returns and standard deviation of returns, thus obtaining distributions for all of them. We summarize these findings as follows.

Table 9. Distribution of Return Performance Metrics

	Min	First Quantile	Median	Mean	Third Quantile	Max
Mean	-0.002	0.007	0.007	0.007	0.007	0.018
Std	0.003	0.010	0.011	0.014	0.016	0.074
Sharpe Ratio	-0.062	0.413	0.629	0.582	0.746	2.136
Overall Return	-0.198	0.612	0.658	0.634	0.666	2.259
CAGR	-0.036	0.083	0.088	0.085	0.089	0.218

Mean stands for the mean of returns that each individual realizes. Similarly, Std. Dev. stands for the standard deviations of of returns obtained by each individual. For instance, an element of Mean distribution is 0.03. This means that the corresponding individual's average of the realized returns is 0.03. Same logic applies for the standard deviation distribution. Overall Return stands for each investor's total return over the course of 6 years. In the first two columns we observe that individual investors have low means of monthly returns accompanied by very low standard deviations which pave the way for relatively high Sharpe ratios. We see that average overall return is 0.63 for our sample which corresponds to 0.09 compound annual growth rate. This average overall return is considered to be very low when we think of the inflation in that era averages to 8 % . It can clearly be seen that the average pension fund investor do not take any risk by investing in equities, thus suffers from low realized returns. Note that it is crucial for investors to time the market where the equity risk premium may fail to exist (depending on the time interval.). In the next chapter we investigate if there is any predictability- and an implied, automated trading strategy- in equity returns that would create economic significance and improve the performances of the portfolios of pension fund investors by removing the behavioral biases considered in this chapter.

CHAPTER 3

RETURN PREDICTABILITY

We start by stating that there is no to very little apparent equity risk premium in Turkey. We obtained Bloomberg Turkey Local Currency Bond Index and BIST100 index in TL from Bloomberg to use in Mann Whitney Wilcoxon Rank Sum Test. We use a historical method to estimate ex-post equity risk premium. Note that since our estimation is historical it is prone to criticisms that historical ERP estimation methods bare. One of the biggest criticisms is that the historical window used for estimation is changing the result of the estimation. For example when we include the post sub-prime crisis era where the recovering of most equity markets occur, we can see a significant ERP. However if we exclude it, we find very little ERP. Table 10 presents the result of the Mann Whitney Wilcoxon Rank Sum Test.

Table 10. Mann Whitney Test For Bonds and Equities

W	p-value
346740	0.6231

The p-value of the test is 0.6231. We can not conclude that returns from equities in Turkey is significantly different from returns gained from Bonds.

We move on to the issue of equity return predictability. The following sections of this chapter is organized as follows. In Section 3.1 we describe the data we used. In section 3.2, we employ an in-sample return prediction model which is widely used in the literature. In section 3.3, we use out-of-sample rolling regressions in order to make meaningful forecasts. In section 3.4, we develop a trading strategy to extract economic significance of the return forecasts. In section 3.5, we compare the performance of the trading strategy and the performances obtained investors.

3.1 Data

We use different sources of data in the analysis. Most of the data is obtained through Bloomberg Professional Services. For stock market return we use BIST100 index returns, which is available between the dates 1988-2017. However in order to avoid structural changes in the data we use the data starting after 2003. Dividend Yield and Price/Earning variables are obtained from Datastream because of the data availability issues. The most popular indicator variables through this thesis are Price/Earnings Ratio and Dividend Yield. The method used by Datastream in order to approximate dividend yield and price/earnings ratio can be found in Appendix A.

3.2 In-sample predictive regression models

There is some evidence that the returns are predictable. Some scholars assert that this is some sort of market inefficiency, some assert that predictability is a rational change over time in the expectations. For example, during or immediately after the 2008 crisis the long horizon expected returns are shown to be high. However, this is a result of the perceived risk by the investors rather than a lucrative secret that a little people know.

We use a simple regression framework using Dividend Yield and Price/Earnings Ratios as explanatory variables. The predictive power of such variables are famous among the literature. Fama and French (1989) for example links this relation to business cycles. In Figure 1 you can see a plot of Dividend Yield versus Forward Returns(returns that we are trying to predict.). Note that the axes are different, but the correlation is indisputable. High dividend yields mean low price levels which mean high future returns and vice versa.

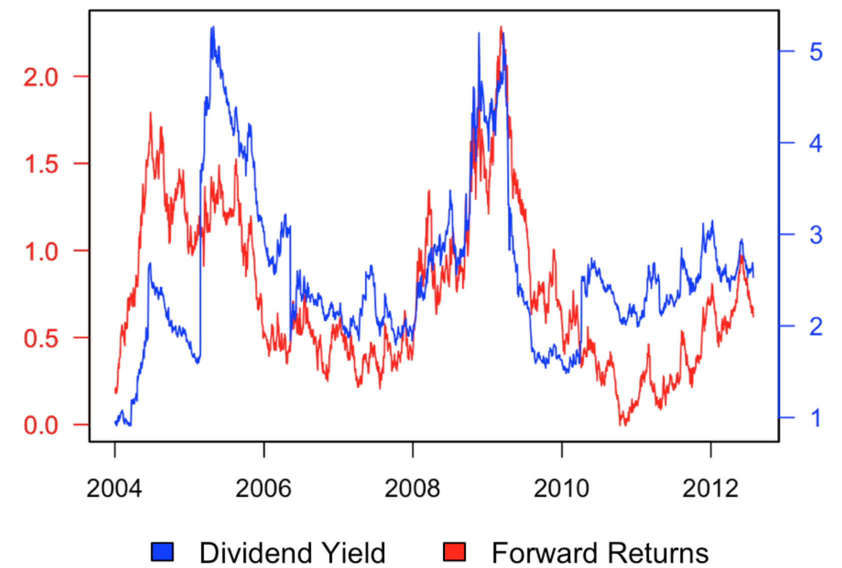

Figure 1. Dividend yield vs 5 year forward returns with time axis

The close relation of forward returns and dividend yield is not so unique. Generally speaking, any variable rather than dividends that detrends price can be used in order to predict returns. Another way to put this is that dividends and prices are co-integrated. So if one of these variables behaves unusually, price of the stock adjusts. Thus expected forward returns increase.

The regression models used in this section is as follows:

$$R_{t:t+h} = \alpha + \beta I_t + \epsilon_t \ ,$$

where h is the investment horizon, I_t is the Indicator Variable which will be used as several different variables like Dividend Yield D_t/P_t or Price/Earnings Ratio.

In Table 11 you can see two different in sample estimations. In the first model the dependent variable is 5 years forward return where independent variable is the dividend yield. In the second one , dependent variable is 1 week forward return and the independent variable is same. Notice that we have significant βs for each model. In the first model which uses long horizon returns for estimation we see a significantly higher β. In addition the R^2 is also higher. This is the exact phenomenon which is prevalent in return predictability literature, which answers the question ”Are returns predictable ?”. According to our analysis and the literature the answer is ”In the long horizons they are, wheras in the short horizons they are not. ”

Table 11. Regression Summary

	Dependent Variable:	
	Forward Returns	
	(1)	(2)
Dividend Yield	0.290^{***}	0.003^{***}
	(0.009)	(0.001)
Constant	0.005	-0.004^{*}
	(0.026)	(0.002)
Observations	2,159	3,454
R^2	0.307	0.003
Adjusted R^2	0.307	0.003
Residual Std. Error	0.381 (df = 2157)	0.038 (df = 3452)
F Statistic	956.479^{***} (df = 1; 2157)	10.464^{***} (df = 1; 3452)

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

In Figure 2 and Figure 2 we see that as we decrease the forecast horizon (h) we linear relationship seems to disappear, which is an interesting phenomenon. Cochrane (Appendix B) shows that this phenomenon is a result of the fact that Dividend Yield- or any other indicator variable- is highly persistent. Similar arguments can be made if we use Price/Earnings Ratio. The regression is summarized in Table 12.

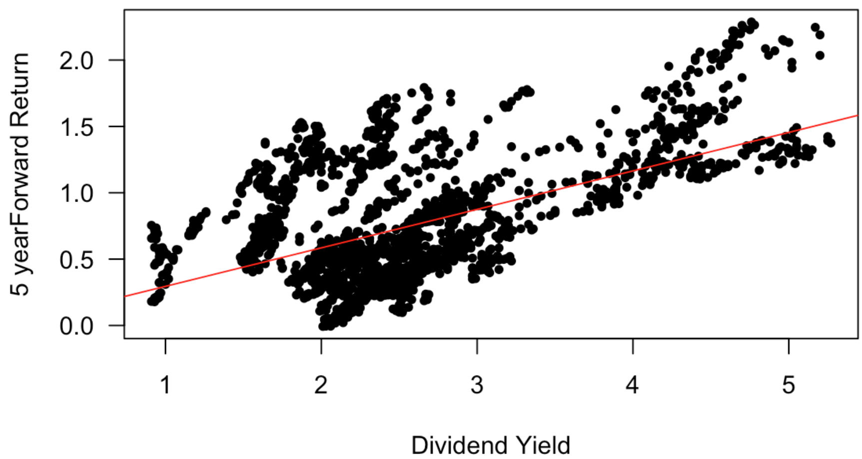

Figure 2. Dividend yield vs 5 year forward returns

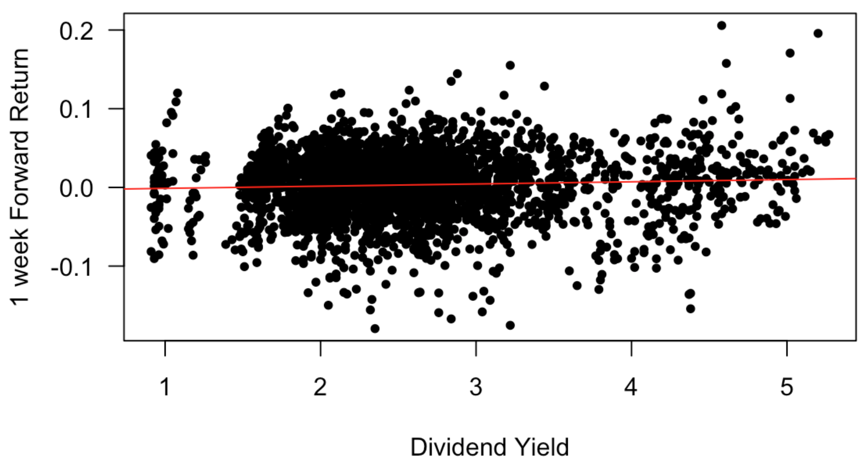

Figure 3. Dividend yield vs 1 week forward returns

Figures 4 and 5 shows the long horizon predictability phenomenon for the Price/Earnings Ratio indicator variable.

Table 12. Results of Price/Earnings Ratio

	Dependent variable:	
	Forward Returns	
	(1)	(2)
Price Earnings Ratio	-0.051^{***}	-0.0005^{*}
	(0.003)	(0.0002)
Constant	1.237^{***}	0.007^{***}
	(0.032)	(0.003)
Observations	1,656	2,951
R^2	0.189	0.001
Adjusted R^2	0.188	0.001
Residual Std. Error	0.392 (df = 1654)	0.038 (df = 2949)
F Statistic	384.459^{***} (df = 1; 1654)	3.574^{*} (df = 1; 2949)

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

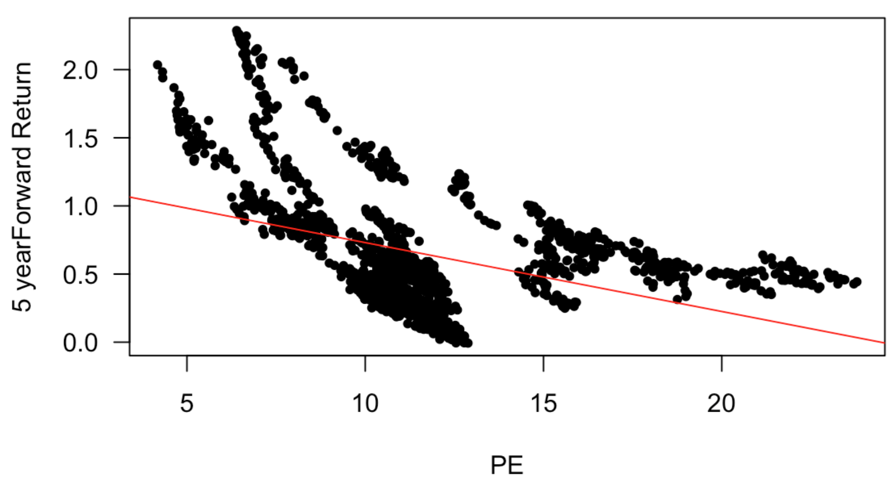

Figure 4. Price/earnings ratio vs 5 year forward returns

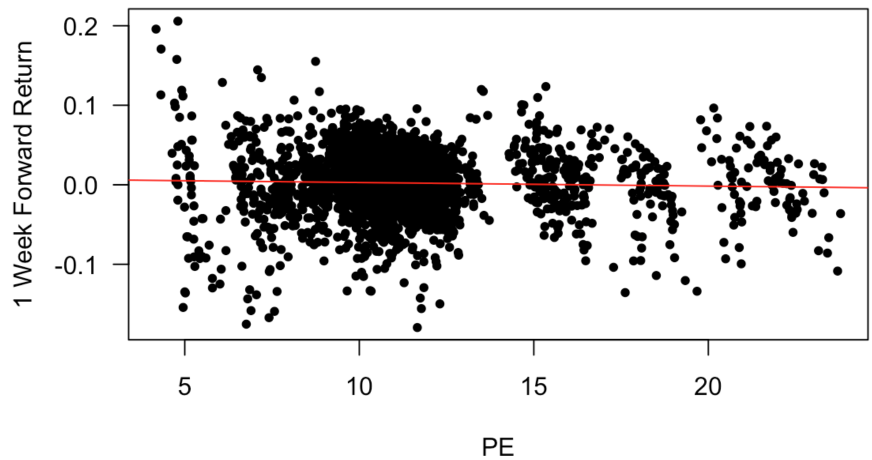

Figure 5. Price/earnings ratio vs 1 week forward returns

3.3 Out-of-sample predictive regression models

In the previous chapter we used in-sample regression models, in which we used data that is not occured during the regression. For example say we are in 2013-01-01 ($t = \tau$) and we want to do the same analysis. If we fit D_t/P_t to R_{t+5y} between the dates, say 2005-01-01/2010-01-01, we would encounter with a serious problem which is that the data points for the forward returns between 2008-01-01/ 2010-01-01 would not been occurred until the date 2015-01-01.

This concept and the out of sample forecast model that we employed can be better understood via Figure 6. In Figure 6 ”Data Lost I” is the data lost due to the fact that we use forward returns. We use ”in sample data” in order to forecast the $R_{t:t+h}$. This procedure has its own down side because as we increase the forecast horizon which increases the predictability(see section 2.2) we get further away from the time τ.

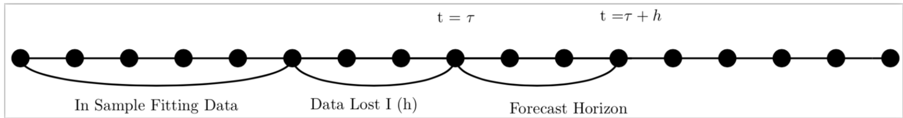

Figure 6. Out of sample estimation procedure

In this section we use a procedure that is as follows:

Algorithm 1 Forecast $R_{t:t+h}$ for $t = \tau, \tau + 1...0 - h$

for $t = \tau, \tau + 1...0 - h$ **do**
  Fit the model using in-sample data $R_{t:t+h} = \alpha + \beta I_t + \epsilon_{t+1}$,
  Forecast $\hat{R}_{t+h}$
end for

We define $OOS - R^2$ as the main performance metric that we want to maximize,we do so because it is hard to report the coefficients for every regression since we fit the model in every time step. The $OOS - R^2$ metrics obtained through the procedure described are as follows. Note that we were not able to run the regressions for 5 year forward returns since we lose extra 5 year worth of data as described in Figure 6. Results shows a similar pattern to those we observed in section 3.2. As we increase the forecast horizon the $OOS - R^2$ increases significantly. In addition, we tried a bunch of variables as Indicator Variables which did not work well. We used EMBI Global Spread to use as a proxy for risk. You can see the results in Table 13. It is crucial to note that the data selection plays a very key role here. A similar argument is made when we try to estimate Equity Risk Premium. Especially, If there are structural changes in the data, data selection plays a key role.

Table 13. OOS-R^2's of Rolling OOS Forecasts

Indicator	1 Week	1 Year	3 Years
Dividend Yield	0.02	0.59	0.80
Price/Earnings Ratio	0.04	0.66	0.62
EMBI Global Spread	0.02	0.12	0.67

3.4 Trading strategy

In this section we develop a trading strategy, implied by the regressions. Note that vast amount of trading strategies can be derived using different predictors. We employ a trivial yet effective trading strategy to test whether an average investor can extract economic significance by the automated nature of the trading algorithm. We also present the performance metrics of a simple buy and hold strategy. Note that we do not use the forecasts itself, rather we use directly the predictor variables. The main reason is the data availability issues described in section 2.3. For instance if we want to predict the 5 year forward return in the first month of 2008 and use 5 years of data for fitting we would need the indicator variables to start from 1998. Even if data were available the severe structural change in the beginning of 2000's would lead unvaluable forecasts as described in section 3.2. So we employ an implied trading strategy by directly using the predictor variables. The summary of the performance metrics are presented in Table 14.

The algorithm that makes the position decisions are as follows.

Algorithm 2 Take position according to indicator variables

for Every time step : $t = 2008 - 01, .., 2013 - 12$ **do**
  Select indicator variables' last 5 year of data : $PE_{t,t-5year}$ and $DY_{t,t-5year}$
  Find z scores of indicator variables at time t : Z_{PE_t}
  Treshold the z-scores
  Compute the average long or short position implied in the last 20 days : $position \in \{-1, +1\}$
end for

Table 14. Performance Metrics of the Trading Strategy

Performance Metrics	Trading Strategy	Buy-Hold
Mean	0.014	0.015
Standard Deviation	0.076	0.075
Sharpe Ratio	0.194	0.202
Overall Return	1.583	0.778
CAGR	0.165	0.100

3.5 Comparison of performances

When we look at the algorithm's performance we see that the mean of the monthly returns is 0.014 whereas an average investor's mean of monthly returns was 0.007. The returns obtained by the algorithm's standard deviation is much higher, (0.076 where investors' is 0.014) as expected since the equity allocation of investors are very low. This makes the Sharpe Ratio for an average investor to be 0.58 whereas the trading strategy obtains a Sharpe Ratio of 0.19. The overall return of the strategy is 1.583 over the course of 6 years which corresponds to a compound annual growth rate of 16.5 % which is approximately double of an average investor and almost greater than of 99 % of the individuals. Even though the Sharpe ratio is low, 172988 people out of 17298 would be better off if they were to use this trading strategy.A simple buy and hold strategy has a Sharpe ratio of 0.20 whereas the overall return (and thus CAGR) is much lower that that of trading strategy but higher than the average investor. In that sense we can confirm that average investor is a holding sub-obtimal level of equities in her portfolio. Note that allocating 100 % of the portfolio is not optimal, nevertheless significant improvements in the fund performance can be achieved by allocating a fraction of portfolio to equities.

CHAPTER 4

CONCLUSION

In the Portfolio Choice and Trading we investigated the equity allocation and trade decisions, which are prone to behavioral biases (equity allocation and trading decisions are affected by demographic characteristics, serial correlation in these decisions, people do not take any risk by holding equities.). We observed that most of the individuals inefficiently do not include equities in their portfolio. We can better observe the implications of this phenomenon by looking at the individual investors return performances realized in the pension plan. The average compound annual growth rate of an average individual investor is 8.5 % which is unquestionably unsatisfactory performance for a pension fund, considering the fact that average inflation in that era is 8%. Note that even a simple buy and hold strategy in equities can upgrade the performance of individual investors. Moreover, We see that several demographic characteristics influence these decisions. For example, being male and married significantly increases the probability of making a trade in the portfolio or being married decreases the amount of average equity holdings. Additionally, when we look at the cross and auto-correlations of returns and decisions(change in equity allocation and making a trade) we see that there are no cross-correlations between returns and portfolio decisions but there is significant positive serial correlation in making change in equity proportions in portfolios.

In the Return Predictability chapter we obtained very similar results to the literature. We observed a little predictability in the short run however as we increase the return horizon the predictability also increased. Especially in the out of sample forecast section we see very promising forecast performances with high $OOS - R^2$. We see that even a simple trading strategy on equities implied by the results of this analysis , would be very useful for the individuals who lack the required financial intellectuality for making such decisions. We obtained a compound annual growth rate of 16.5 % by implementing a simple yet effective automated trading strategy whereas the simple buy& hold strategy obtains a CAGR of 10%. Systems that are already present in developed markets like robo-advisory platforms could better off the individual investors by also making the financial systems more healthy by increasing the participation rate of equity markets. Note also that such systems would remove the behavioral biases - including concepts like panic selling or even moral hazard- which jeopardize the performance of individual investors.

APPENDIX A

P/E AND DIVIDEND YIELD CALCULATIONS BY DATASTREAM

Price/Earnings ratio is calculated by Datastream as follows:

$$PE_t = \frac{\sum_1^n (P_t * N_t)}{\sum_1^n (E_t * N_t)}$$

where;

PE_t = Price/Earnings ratio at day t

P_t = Unadjusted share price at day t

N_t = Number of share on issue at day t

E_t = Earnings per share on day t

n = number of constituents in index

Dividend Yield for BIST100 is calculated by calculating the total amount of dividend and expressing this quantity as a fraction of total market value for constituents of that sector.

$$DY_t = \frac{\sum_1^n (D_t * N_t)}{\sum_1^n (P_t * N_t)} * 100$$

where;

DY_t = Aggregate Dividend Yield at day t

D_t = Dividend per share on day t

N_t = Number of share on issue at day t

P_t = Unadjusted share price at day t

n = number of constituents in index

APPENDIX B

LONG HORIZON PREDICTABILITY

In his lecture notes Cochrane states that long horizon predictability when a slow varying (highly persistent) indicator variable is used, follows from mathematically from short horizons. Let x_t denote this persistent indicator and r_t log returns .

$x_{t+1} = \phi_1 x_t + \varepsilon_t$

$r_{t+1} = \beta x_t + \epsilon_t$

Using this two regressions to write a longer horizon return regressions.

$r_{t+1} + r_{t+2} = (\beta x_t + \epsilon_t) + (\beta x_{t+1} + \epsilon_{t+1})$

$r_{t+1} + r_{t+2} = \beta(1 + \phi_1)x_t + \beta\varepsilon_{t+1} + \epsilon_{t+1} + \epsilon_{t+2}$

Similarly;

$r_{t+1} + r_{t+2} + r_{t+3} = \beta(1 + \phi_1 + \phi_1^2)x_t + (\beta\varepsilon_{t+1} + \beta\varepsilon_{t+2} + \epsilon_{t+1} + \epsilon_{t+2} + \epsilon_{t+3})$

Thus it as the horizon increases the coefficient estimated increases as well.

REFERENCES

Agnew, J., P. Balduzzi, & A. Sunden, (2003). Portfolio choice and trading in a large 401k plan *American Economic Review*, 93 (1), 193-215.

Barber, Brad M., & Odean, Terrance. (2000). Trading is hazardous to your wealth: The common stock investment performance of individual investors. *Journal of Finance* , 55(2),773806.

Campbell, J., & Shiller, R. (1988). The dividend-price ratio, expectations of future dividends and discount factors. *Review of Financial Studies*, 1, 195227.

Cenesizoglu, T., & Timmermann, A., (2012). Do return prediction models add economic value?. *Journal of Banking & Finance* , 36(11), pages 2974-2987.

Cochrane, J., (1999). New facts in finance. *Economic Perspectives, Federal Reserve Bank of Chicago*, 23, 36-58.

Cochrane, J., (2008). The dog that did not bark: A defense of return predictability.*Review of Financial Studies*, 21(4), 1533-1575.

Donadelli, M., & Prosperi, L., (2012). *The equity risk premium: Empirical evidence from emerging markets.* (Report No : 1201), Rome, Italy : CASMEF Dipartimento di Economia e Finanza, LUISS Guido Carli.

Fama, E. & French, K., (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics*, 25, 2349.

Ferson, W. & Harvey, C. (1993). The risk and predictability of international equity returns. *Review of Financial Studies*, 6, 527566.

Henningsen , A. (2012). Estimating censored regression models in R using the censreg package. http://CRAN.R-project.org/package=censReg.

Jagannathan, Ravi, & Kocherlakota, Narayana R. (1996). Why should older people invest less in stocks than younger people? *Quarterly Review, Federal Reserve Bank of Minneapolis*, 2(3), 1123.

Odean, Terrance. (1999). Do investors trade too much? *American Economic Review*, 89(5), 127898.

Salomons, R. & Grootveld, H., (2002). *The equity risk premium: emerging versus developed markets.*(Report No: 02E45.), Groningen, Netherlands : SOM Institute, University of Groningen

Welch, I., & Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21, 14551508.