IFA
Iranian Journal of Finance
Print ISSN 2676-6337
Online ISSN 2676-6345

Comparison of linear regression models Ordinary Lasso, Adaptive Group Lasso and Ordinary Least Squares models in selecting effective characteristics to predict the expected return

Raheleh ossadat Mortazavi
Department of financial management, Kish International Branch, Islamic Azad University, Kish Island, Iran. (Email: mortazaviraheleh@yahoo.com)

Hamid Reza Vakilifard*
*Corresponding author, Associate Prof., Department of management, Islamic Azad University, Science and Research Branch, Tehran, Iran. (Email: Vakilifard.phd@gmail.com)

Ghodratallah Talebnia
Associate Prof., Department of Accounting, Islamic Azad University, Science and Research Branch, Tehran, Iran.(Email: gh_talebnia@yahoo.com)

Seyedeh Mahboobeh Jafari
Assistant Prof., Department of Economic and Accounting, Islamic Azad University, South Tehran Branch, Tehran, Iran. (Email: Jafari.mahboobeh@gmail.com)

Abstract
In this study, for the selection of the characteristics of the company that provides the incremental information to investors and financial analysts, the linear models are adapted by the ordinary Lasso method (Tibshirani, 1996), Adaptive Group LASSO (Zu, 2006) and the least squares method (OLS). The main objective of this research is to determine which method can predict the expected return on stock portfolios in the shortest time and using the least effective features. The research sample is1340observations, including 134companies listed in Tehran Stock Exchange, and the research variables from the financial statements of the companies and the stock market reports between 2008and 2018. The results of this study show that by employing the least squares regression method, 7 characteristics, the typical 5- characteristics LASSO method and in the Adaptive Group LASSO method, only 4characteristics, contain incremental information to predict the expected returns of stock portfolios. In the second place, by applying the Adaptive Group LASSO regression method, one can achieve the same results with using the least characteristics.

Keywords: LASSO Regression, Adaptive group LASSO Regression, Ordinary Least Squares Regression, Expected Returns of Portfolios.

DOI:

Introduction

Prediction of expected return on portfolio is an important criterion in investment, portfolio allocation and cross-sectional analysis. According to the importance of this issue, financial theorists have proposed and investigated many features from the past to accurately estimate expected returns. Of course, it should be borne in mind that not necessarily any variable affecting returns is capable of enhancing equity pricing models in determining expected returns. Variables as "factors" are incorporated into these models that can explain the simultaneous and joint changes in the returns of a large group of companies [31]. Efficiency models are divided into two groups of single-factor models and multi-factor models. The basic concept in the single-factor model is that all securities are affected by market fluctuations, as similar economic forces will affect most companies in the future. One-factor models include the Capital Assets Pricing Model and the market model. Gradually, the use of multi-factor models to explain stock returns replaced the single-factor model of Capital Assets Pricing Model. Since the 1980s, researchers have been trying to find a relationship between variables other than beta to predict stock returns and have achieved some success, including the ability to measure earnings per share (BASU, 1977), firm size variable (Benz, 1981), book value to stock market value (Rosenberg et al., 1985), past stock returns (De Bonnet &Thaler, 1985), leverage (Behandari, 1988), profitability (Hagen & Baker 1996). Many studies have been done to describe the cross-section of expected returns. In some studies, such as Rubinstein(1976), Lucas (1978) and Brindle (1979), fewer variables and in some, such as Harvey (2016), examined hundreds of articles and features that had greater predictive power for the expected cross-section of returns. [26] Which of these factors and characteristics have the greatest impact on predicting returns and providing incremental information to the researcher and investor is importance of these studies.

Although these models have very valuable applications, they are not capable of answering Cochran's (2011) question as to: Which characteristics really provide “independent information about average return?Researchers typically use two methods to identify predictors of performance: 1- portfolio sorts based on one or more characteristics (including: beta, company size, book value to market value, etc.) 2- linear regression in spirit of Fama and Macbeth (1973).According to Cochran (2011), with regard to the introduction and evaluation of new characteristics, different methods should be used to identify and rank effective features in predicting expected returns [13].

Various studies have been done on the characteristics and factors affecting the expected return on stocks of listed companies in Tehran Stock Exchange, while the results of these studies indicate differences and even contradictions of the findings of these studies. The purpose of the present study is to determine which of the linear regression methods including ordinary Lasso, adaptive group Lasso and ordinary least squares methods in the least time and using least characteristics provide incremental information to predict expected returns.

Literature Review

Factor models state that stock or portfolio returns are influenced by various factors, thereby estimating expected returns. Since the introduction of the beta factor or characteristic in the one-asset pricing model, so far many features have been presented and studied by financial researchers. As a result, given the enormous volume of financial data being produced and the large amount of observations, variables, time, and cost involved in analyzing it, it seems necessary to simplify the issues in data analysis and research. In other words, we can get better results with fewer variables. Lasso's regression is useful in linear multivariate analyses. Using Lasso regression, we try to provide a suitable method for modeling the response variable based on the least and of course the most appropriate number of independent variables. This method seeks to isolate more appropriate variables than other variables and provide a simpler model.

In the empirical test of the capital asset pricing model, Fisher et al. (1974) observed a linear relationship between portfolio returns and their beta. They studied the stock price trend of New York Stock Exchange companies. The study found that the relationship between risk and return is linear [19]. Fama and French (1992) summarize the findings of previous empirical studies using the Fama and Macbeth cross-sectional regression method, the relationship between beta variables, firm size, book-to-market ratio, financial leverage, and profit-to-price ratio with expected returns. Studied the stocks in the US capital market and concluded that systematic risk (beta) did not have the power to explain the difference in stock returns during the period 1963–1990, and of the variables studied, two variables were “firm size” and “book value to market valueratio” is better able to explain the average return on equity [17].

Busart and Hillin (1999) implemented models to predict excessstock returns based on international data to determine the best predictor variables (such as price-to-earnings ratio, dividends, and short-term interest rates) of excess stock returns. Taken together, they concluded that excess stock returns could be predicted, as well as evidence that out-of-sample predictive ability was zero.

Olsen and Mossman (2003) investigated the prediction of stock returns by using financial ratios. In this study, neural networks model and ordinary least squares method were used to predict stock returns. The results showed that the neural network method hasmore acceptable results than other prediction methods and the prediction error was significantly reduced [38].

Avramov and Cordia (2006) examined Capital Asset Pricing Models, Consumer Capital Asset Pricing, Fama& French, Fama& French plus Pastor and Stamburg's Liquidity Risk, Fama& French plus the winner and loser factors and the Jonathan and Wang model to explain some of the turmoil on the New York Stock Exchange, Nasdaq and Amex. They argued that multi-factor models whose beta changes over time better explain the effects of disruptions such as the size and book value to market value ratio [2].

In an article entitled "Selecting a Portfolio Based on a Financial Strength Index Using Data Coverage Analysis", Edirising and Zhang (2008) used a series of financial ratios to estimate firms' financial strength and correlate these metrics with actual stock returns. The financial ratios used in this study fall into six categories that include profitability measures (including return, capital, return on assets, net profit margin and earnings per share), operating efficiency measures (including accounts receivable, inventory turnover, asset turnover),liquidity measures (including current ratio, quick ratio, debt-to-equity ratio), leverage measures (including leverage ratio, total debt-to-asset ratio, total debt-to-equity ratio), corporate vision criteria (including price-to-equity ratio),revenue and the ratio of market value to book value and growth metrics (including earnings growth rate, net profit growth rate and earnings growth per share rate) [16].

Lee, Eric, and Wang (2011), in a study called “Assessing Implicit Cost Estimates”, estimated expected returns through seven different methods and compared them with expected returns based on beta. This study considers expected beta-based returns equal to the returns from capital asset pricing, Fama and French three-factor models and four-factor models. The results show that actual returns are not a good indicator for predicting expected returns, and the estimates of each of the models under study in this study are more accurate than the beta-based estimates.

Elliott et al in their research (2013) proposed a new model based on combining linear predictions based on subset regressions to predict additional stock returns. The results of this study show that combinations of subset regressions can lead to higher accuracy of predictions than ordinary methods and predictions based on equal weights.

Uddin et al in their study (2013) used regression analysis to examine the factors influencing stock price in the Dhaka Stock Exchange. The empirical results of this study indicate that the earnings per share, net asset value, net profit after tax and price to income ratios are the most important variables affecting the stock price in the country [49].

According to Lewellen's (2015), the study of cross-sectional forecasts of return on assets using Fama Macbeth regression, especially in the short run, estimates the expected return better. These predictions simulate how an investor in fact combines multiple characteristics of a company to obtain an estimate of the expected return on equity. Lewellenet al. (2015) jointly studied the predictive power of fifteen characteristics and found that only a few of them are important predictors of expected cross-sectional returns [32].

Green et al. (2016) validated Lewellen'sresults for more features and fewer samples in 1980 and confirmed his main result [24].Chinco, Clark, and Ye (2015) considered unreachable conditions in Menschauen and Buhlmann's (2006) research to achieve fixed model selection in one-step Lasso. They applied a linear model to predict high repeatability returns using past earnings of their respective stocks and found that their model increases the predictive capability of ordinary least squares [11].

Recent applications of Lasso methods in finance have been by Huang and Shay (2016). They used anadaptive group LASSO in a linear model and applied macro factors to test for determinants of bond risk premia [28].Brizgalova (2016) noted in a study of "incorrect factors in linear asset pricing models" that poor identification in linear operating models can lead to overestimation of significant cross-sectional risk factors. She proposed a shrinkage-based estimator todetectpossible rank deficiency in the design matrix and to identify strong asset pricing factors [10].

Giglio and Chiao (2016) proposed a three-pass regression method that combines principal component analysis and a two-stage regression to estimate consistent factor risk premia in the presence of omitted factors when the cross section of test assets is large[23].Light et al. (2016) used partial least squares to summarize the predictive power of firm characteristics for expected returns. Partial least squares summarize the predictive power of all characteristics and thus does not directly distinguish important characteristics and does not reduce the number of characteristics for predicting returns [33].

Sheari(2004) examined the role of fundamental accounting information in predicting stock returns. He used Stuart's deductive model to select accounting variables related to returns and then expanded on its constituents. The results show the predictive power of accounting information [45].Shah Nazari (2005), in a research on alternative systematic risk metrics, using multivariate regression, concluded that according to capital asset pricing model (assuming stability of other variables), beta has weak power in explaining returns. But in Fama and French multivariate models, beta has a significant relationship with returns, along with other variables [46].

Namazi and Rostami (2006) examined the relationship between financial ratios and the rate of return on listed companies in Tehran Stock Exchange and concluded that there was a significant relationship between financial ratios and the rate of return on equity [35].Zamazi (2010) investigated the relationship between cash flow-based financial ratios and equity returns in Tehran Stock Exchange. According to this study, financial ratios (current debt coverage power ratio, interest coverage power ratio, cash flow per share and price-to-equity ratio) were studied. Cash flows per share have no significant relationship with stock returns [50].

Beshkooh and Afshari(2012) evaluate the factors affecting the stock market by using hierarchical analysis. The results of this study show that stock price, dividend, earnings per share, company management, operating profit ratio, technology, price to earnings ratio, company size, economic value added and systematic risk have higher share and weight than other variables are influencing the stock market [7].Janani et al. (2012) in a research using TOPSIS technique investigate factors affecting portfolio selection in Tehran Stock Exchange over a five year period. The findings of this study indicate that the variables of dividend ratio, systematic risk, volume of transactions, and price-to-earnings ratio are influential variables in portfolio selection.[29]

Pourzamani and Ali Bashiri (2013) examined the performance of the four-factor Carhart model for predicting expected returns by dividing growth and value stocks. In his research by using multivariate regression and paired t-test, he found that stocks were more profitable in emerging markets in Iran [42].

Research methodology

Past studies suggest that many characteristics and factors have been used to predict expected returns. By examining the theoretical foundations and background of the study, it can be seen that financial theorists have failed to answer which of the firm's characteristics provide more useful and effective information for predicting the expected return on the portfolio for financial analysts and investors. In this study, based on past studies on corporate characteristics and their impact on predicting expected returns, 36 features were used to design the model. The following model is the conceptual and theoretical model of research.

In order to calculate the portfolio's incremental return (independent variable), portfolio returns were compared with the Tehran Stock Exchange's TEDPIX Index.

The return on investment of a portfolio that performs better than a benchmark or index with a similar level of risk is called excess return. This return is widely used to measure the surplus value created by the portfolio or investment manager, or to measure the ability of management to overcome the market. Another name is Alpha Excess Returns.

- The following equation is used to calculate the excess return:

$$\text{ER (Excess Retrn)} = R_{it} - R_{mt} \tag{1}$$

- The following equation is used to calculate R_{it} stock returns:

$$R_{it} = \frac{(1+a+\beta)P_1 - P_0 - 1000\beta}{P_0} \times 100$$

R_{it}: The actual return of company i in period t
P1: Stock prices at end of period (2)
P0: Stock prices at the beginning of the period

- The percentage of total capital increase from the place of reserves and accumulated profits

Percentage of capital increase from cash and receivables

The nominal price per share of stock companies is 1000 Rials

The following equation is used to calculate the R_{mt} TEDPIX index return:

$$R_{mt} = \frac{P_1 - P_0}{P_0} \times 100$$

R_{mt}: Total index returns in period t
P1: Market index at the end of the period (3)
P0: Market index at the beginning of the period

The independent variables that comprise the 36 characteristics that influence return prediction are described in Table 1.

Table 1. describes the independent variables used in the research

o	Featurs	Abbreviation	calculation method	Studied by
	Returns 2 to 1 month ago	r2-1	Short-term reverse investment, returns a month earlier	Jegadeesh(1990)[30]
	Returns 12 to 2 months ago	r12-2	Total returns from 12 months to 2 months beforethe return prediction	Fama& French(1996)[18]
	Returns 12 to 7 months ago	r12-7	Total Returns 12 months to 7 months beforethe return prediction	Novy-Marx(2012)[38]
	Returns 36 to 13 months ago	r36-13	Long-term reverse investment, Total Returns 36 months to 13 months beforethe return prediction	De Bondt&Thaler(1985)[15]
	Investment (asset growth rate)	investment	Percentage of annual growth rate of total assets	Cooper, Gulen&Schill(2008)[14]
	Change in property, plant	PI2A	Changes in property, machinery, equipment and inventory on	Liandrs, Sun & Zhang(2008)[35]

	and equipment		totalassets	
	Percentage change in stock	Shrout	Percentage change in published shares	Pontif& Woodgate(2008)[41]
	Changes of the book value of equity	CEQ	Percentage change in the book value of equity	Richardson et al.(2005)[44]
	Sales to cash	S2C	Net sales to total cash and short-term investments	Ou& Penman(1989)[40]
0	Circulation of capital	CTO	Net sales to total assets	Haugen & Baker(1996)[27]
1	Earnings per share	EPS	The proportion of pre-invoiced earnings to published shares	Basu(1983)[6]
2	Marginal profit	PCM	The result of the difference between the net sales and cost of goods sold divided by net sales	Bustamante &Donangelo(2016)
3	profit margin	PM	Operating profit after depreciation on sales	Soliman(2008)[48]
4	Profitability	PROF	Gross profits divided by book value of equity	Ball, Gerakos, Linnainmaa&Nikolaev(2015)
5	Asset turnover	SAT	The ratio of sales to assets	Soliman(2008)[48]
6	Asset returns	ROA	Profit before extraordinaryitems to total assets.	Balakrishnan, Bartov&Faurel(2010)[3]
7	Return on equity	ROE	Income before extraordinary items to the book value of equity	Haunge& Baker(1996)[22]
8	Absolute operational Accrual	AOA	Absolute of Changes in noncash working capital minusdepreciation scaled by total assets	Badyopadhyay , Huang &Wirjanto(2010)[5]
9	Operational leverage	OL	Total cost and sales, general and administrative expenses on total assets	Novy-Marx(2011)[37]
0	Tangibility	Tan	0.715% of total receivable accounts plus 0.546%ofgoods inventories plus 0.535% of property, machinery and equipment ,plus cash and short-term investments divided by total assets.	Hahn & Lee(2009)[25]
1	Operational Accruals	OA	Changes in noncash working capital minus depreciation scaled by total assets	Sloan(1996)[49]

2	Assets to market investment	A2ME	Total assets on market investments at the beginning of the period	Bhandari(1988)[8]
3	book value to market value	BEME	The book value of equity is equal to the total assets minus total debt. The value of the equity market is multiplied by the number ofshares published at the .۱market price per share in t-	Rosenberf,Reid&Lanstein(1985) & Davis, Fama& French(2000)[45]
4	Cash ratio	C	The ratio of cash and short-term investments to total assets.	Palazzo (2012)[42]
5	cash to the price	C2D	The ratio of income and expense items and depreciation of tangible and intangible assets to total debt.	-
6	profit to price	E2P	Ratio of income before extraordinary items to the market capitalization at t-1	Basu(1983)[6]
7	Annual sales growth rate	Sales-G	Annual percentage sales growth rate	Lakonishok, Shleifer,&Vishny(1994)[31]
8	Tobin Q	Q	Equity market value minus cash and short-term investments minus deferred taxes scaled to total assets.	-
9	Sales to the price	S2P	Net sales ratio to market investment.	Lewellen (2015)[33]
0	Total assets	AT	Total balance sheet assets	Gandhi &Lustig(2015)[21]
1	Total fluctuations	Total Vol	Standard deviation of excess returns	Ang, Hodrick, Xing & Zhang(2006)[1]
2	SD of daily trading volume	Std-Volume	Standard deviation of daily trading volume	Chordia, Subrahmanyam&Anshuman(2001)[12]
3	Maximum daily return	Ret-Max	Highest daily returns in previous month	Bali, Cakici,& Whitelaw(2011)[4]
4	Beta	Beta	Covariance between stock returns and market returns divided by market return variances	Frazzini& Pedersen (2014)[20]
5	The price to the highest price	R - H price	The ratio of the highest price last month to the highest price of the year	George & Hwang(2004)[22]
6	Leverage	LEVG	Debt to Total Debt and Equity	Lewellen(2015)[33]

The subject area of this research is investment management, portfolio analysis and portfoliomanagement. The sample period of the research is a ten-year period between 2008 and 2018. The data of the research includes companies listed on the Tehran Stock Exchange.

In this study, the listed companies in Tehran Stock Exchange were selected as the statistical population. Required information was extracted from financial statements and other information such as prices, indices and returns, etc. were extractedfrom websites of companies listed on the stock exchange. The reason for this choice is the greater attention of investors and financial analysts to the stock market, the availability and transparency of corporate accounting information. Stock requirements for timely dissemination of accounting information have made the information environment more suitable for research. The statistical population of this research includes companies that have the following characteristics:In order to make the information comparable, their financial year may end in March. They are not part of investment, financial and credit companies and banking services. During the period under study, they have been active in the Tehran Stock Exchange and do not have a change in their financial year during the study period.

In this study, information sources are divided into two categories; the first one deals with the study of theoretical foundations, research literature, and background research using library resources, internal and external journals and databases, articles and theses. The second category is the data collection resources. Since the required information is related to the accounting items contained in the audited financial statements of the companies and the value of the shares, the required data are from the Securities and Exchange Organization Network website, the Comprehensive Publishers Information System at www.codal.ir and the Processing Center Iran Financial Information at www.fipiran.com andManually extracted CDs as well as financial information software including the Rahavard Novin. R and Excel software were used to perform the analysis. In this study, according to the type of data and the statistical analysis method available, panel data were used.

In this study, we seek to answer which of the following methods of ordinary least squares regression, ordinary Lasso regression, and adapted group Lasso regression, employing effective corporate characteristics has the best performance in predicting portfolio returns. Regression functions are widely used to establish a causal relationship between a dependent variable and independent variables. Linear regressions help to study the predictive power of expected returns based on a large number of features in common.

The application of the panel linear regression model to calculate the surplus return of the properties is as follows:

$$R_{it} = \alpha + \sum_{s=1}^{S} \beta_s C_{s,it-1} + \varepsilon_{it}.$$

Whereas; (4)

Rit: Excess return, C: firm's characteristics, i: Firm I, t: time, S: Number of firm's characteristics, α &β : Coefficients and ε_{it}: Error term

For linear regression models, the ordinary least squares method is the simplest and most common method. The original design of this method, usually illustrated by OLS, was put forward by the famous German mathematician Carl Friedrich Gus in the eighteenth century. The idea of the ordinary least squares method is to obtain the model coefficients of values that are closest to the sample regression model for the observations Y1, ..., YT. In other words, show the least deviation from the above observations. The criterion for the ordinary least squares method is that the coefficients must be estimated to minimize the sum of the squares of the residuals. The ordinary least squares method for estimating the coefficients does not require any conditions on the sentence but it is necessary to establish classical assumptions for the coefficients to be unbiased and statistical inference to be possible. The ordinary least squares method as follows;

$$Min \quad \sum_{t} \sum_{i} \left(R_{it} - \alpha - \sum_{s=1}^{S} \beta_s C_{s,it-1} \right)^2$$

Whereas; (5)

Rit : Excess return, C: firm's characteristics, i: Firm I, t: time, S: Number of firm's characteristics, α & β : Coefficients and εit: Error term

Many nonparametric regression methods do not perform well when the number of independent variables is large, and the data scatter in this set causes large estimates of variance to be unacceptable, unless the sample size is extremely large. Interpretability is another nonparametric regression problem based on the kernel and smoothing of SP line estimates. The information of these estimators includes the relationship between independent and dependent variables that are often difficult to understand.

To address these problems, Stone (1985) proposed collectible models. These models estimate an increasing approximation of the multivariate regression function.The benefits of an incremental approximation are at least two:

1- Each of the collectable terms is estimated using a unique variable filter.

۲-The unique criteria explain how the dependent variable is estimated with the independent variables.

In the development of the collectivist model toward a field of distributed families, Hosti and Tibshirani (1990) proposed generalized collectivist models. These models are able to relate the mean of the dependent variable to a summing device via a linear function.While there is a huge amount of data being produced in today's world, researching and analyzing such a large volume of observations, variables, costs a great deal of time and money. So simplifying the issues that appear to be complex is essential in analyzing data. In other words, the results of a better analysis can be obtained with fewer variables. Lasso's regression is useful in linear multivariate analyzes. Using Lasso regression, we try to provide a suitable method for modeling the response variable based on the least and of course the most appropriate number of independent variables. This method seeks to isolate more appropriate variables than other variables and provide a simpler model.

The ordinary Lasso regression method is used for model selection (variable) and parameter estimation simultaneously in regression models. In this method, in order to estimate the β regression coefficients, the sum of the second power residuals with a finite term is minimized, which states that the absolute sum of the coefficients is less than a fixed value. The main feature of the Lasso method is the creation of a thin scattered basket, which means that in selecting the optimal basket it considers a number of assets and excludes the rest of the assets, sets the weights to zero. This method removes those assets that are highly correlated with each other. The objective of LASSO is to solve,

$$Min \quad \sum_{t}\sum_{i}\left(R_{it}-\alpha-\sum_{s=1}^{S}\beta_{s}C_{s,it-1}\right)^{2} \quad subject\ to\ \sum_{s=1}^{S}\left|\beta_{s}\right|\leq t$$

Whereas; (6)

Rit : Excess return, C: firm’s characteristics, i: Firm I, t: time, S: Number of firm’s characteristics, α & β : Coefficients

The ordinary Lasso method assumes the same control parameter for each regression coefficient or, in other words, the same amount of contraction for each regression coefficient, which results the biased estimates and thus the estimators are ineffective and inefficient. For this reason, Zhou (2006) proposed an adaptive Lasso regression method using different control parameters for different regression coefficients with the following objective function:

$$Min \quad \sum_{t}\sum_{i}\left(R_{it}-\alpha-\sum_{s=1}^{S}\beta_{s}C_{s,it-1}\right)^{2} \quad subject\ to\ \sum_{s=1}^{S}w_{s}\left|\beta_{s}\right|\leq t$$

Whereas; (7)

Rit : Excess return, C: firm’s characteristics, i: Firm I, t: time, S: Number of firm’s characteristics, α & β : Coefficients and w: weight vector

In this method, more contraction value is used for the coefficients zero and less contraction value is used for non-zero coefficients, so it is more efficient in estimating the effective parameters and selecting descriptive variables than the ordinary Lasso method.

Research findings

Among the many characteristics used in different pricing and stock return models, 36 characteristics (Table 2-5) were selected for this study. Then, using ordinary Lasso methods, adaptive group Lasso and ordinary least squares method, the selection and identification of factors that provide incremental information for predicting cross-section of expected stock returns is studied. In order to avoid bias in results and uniformity, it is essential that the information of each company be available for a period of ten years. The study period is from 2008 to 2018. Finally, the data of 134 companies listed in Tehran Stock Exchange were used and 1340 observations were analyzed.

The characteristics which are used in this study are detailed in table2.

Table 2. Descriptive statistics of 36 independent variables

	code	**Mean**	**stad.dev**	**q25**	**Med**	**q75**
	r2_1	0.0172	0.2085	-0.06	-0.02	0.06
	r12_2	0.2627	0.8407	-0.16	0.075	0.44
	r12_7	0.1492	0.6569	-0.14	0.02	0.28
	r36_13	0.579	1.4683	-0.07	0.35	0.96
	investment	0.168	0.2623	0.01	0.11	0.25
	PI2A	0.0434	0.1245	-0.02	0.03	0.09
	Shrout	0.3872	1.0647	-0.15	0.07	0.5225
	CEQ	0.1265	2.4327	-0.03	0.1	0.3
	S2C	51.074	115.2998	9.49	20.63	46.765
0	CTO	0.9124	0.714	0.53	0.77	1.06
1	EPS	863.9349	1362.7496	150.89	509.145	1196.1175
2	PCM	0.3089	0.2494	0.15	0.27	0.41
3	PM	5.3914	83.172 5	0.07	0.18	0.33

4	PROF	0.5328	0.9791	0.3	0.505	0.77
5	SAT	0.9248	0.7539	0.53	0.77	1.06
6	ROA	0.14	0.1711	0.04	0.13	0.22
7	ROE	0.2742	2.142	0.14	0.345	0.5425
8	AOA	749646.6171	2527515.999	48434.75	153032	405081.25
9	OL	0.7586	0.6986	0.39	0.59	0.89
0	Tan	0.3555	0.1657	0.25	0.35	0.44
1	OA	-114420.564	2633937.172	-75218.25	34946	194632.11
2	A2ME	1276173.584	1079614.322	557803.965	963483.545	1633144.5
3	BEME	0.5051	0.491	0.24	0.42	0.67
4	C	0.0671	0.0854	0.02	0.04	0.08
5	C2D	0.4482	1.6119	0.06	0.21	0.45
6	E2P	0.1244	0.2028	0.05	0.13	0.21
7	Sales_G	0.2507	1.5131	-0.01	0.14	0.34
8	Q	1.1981	1.0608	0.5	0.89	1.5425
9	S2P	1.3173	1.7753	0.43	0.82	1.61
0	AT	5489716.383	24709828.93	434612.5	962637	2546729
1	TotalVol	0.1762	0.4951	0.06	0.11	0.16
2	StdVolume	1587236.619	7016034.957	68716.9625	259691.965	932660.125

3	RetMax	0.0502	0.1416	0.02	0.04	0.05
4	Beta	0.8209	2.987	0.09	0.51	1.05
5	RHprice	0.7668	0.3108	0.62	0.8	0.92
6	LEVG	-0.0509	2.527	-0.05	0	0.02

Tables (3), (4) & (5) illustrate how many effective characteristics are selected and each of the effective characteristics that are expected to be using each of the linear regression methods examined.

Table 3. Characteristics selection using OLS regression

Firms	Model	penalty	sample	sample size	selected	Characteristics selected						
ALL	linear Model	OLS	FULL	1340	7	r12_2	r12_7	r36_13	CEQ	S2C	Beta	LEVG
r2_1, PI2A, Shrout, CTO, EPS, PCM, PM, PROF, SAT, ROA, ROE, AOA, OL, Tan, OA, A2ME, BEME, C, C2D, E2P, Q, S2P, AT, StdVolume, RetMax are never selected												

Table(3), as can be seen in the ordinary least squares linear regression method, there are 7 characteristics including, 12to 2 months before the return prediction (r12-2), 12 to 7 months before the return prediction (r12-7), 36 to 13 months before the return prediction (r36-13), percentage change in equity book value (CEQ), sales to cash ratio (S2C), beta and leverage (LEVG) are selected as effective characteristics for predicting the expected return.

Table 4. Characteristics selection using LASSO regression

Firms	Model	penalty	sample	sample size	selected	Characteristics				
ALL	Linear Model	LASSO	FULL	1340	5	r12_2	Sales_G	TotalVol	Beta	RHprice
r2_1, PI2A, Shrout, CTO, EPS, PCM, PM, PROF, SAT, ROA, ROE, AOA, OL, Tan, OA, A2ME, BEME, C, C2D, E2P, Q, S2P, AT, StdVolume, RetMax are never selected										

Whereas, according to Table (4) using the ordinary Lasso regression method, 5 characteristics are selected to predict expected returns, including 12 to 2 months before the return prediction (r12-2), annual sales growth rate (Sales-G). , Total Volatility, Beta and price to the highest priceRatio (RHprice) have been identified as effective characteristics in predicting expected returns.

Table 5. Characteristics selection using Adaptive Group LASSO regression

Firms	Model	penalty	sample	sample size	selected	Characteristics			
ALL	Linear Model	A.G. Lasso	FULL	1340	4	r12_2	TotalVol	Beta	RHprice
r2_1, PI2A, Shrout, CTO, EPS, PCM, PM, PROF, SAT, ROA, ROE, AOA, OL, Tan, OA, A2ME, BEME, C, C2D, E2P, Q, S2P, AT, StdVolume, RetMax are never selected									

In the last method used, the adaptive group Lasso linear regression included only 4 characteristics, 12 to 2 months before the return prediction (r12-2), beta, total volatility, and price to the highest priceratio. (RHprice) were selected as the only characteristics capable of predicting expected returns.

Conclusion

Forecasting the expected return on portfolio is an important factor in investment, portfolio allocation and cross-sectional analysis. Given the importance of this issue, financial theorists have proposed and investigated many features from the past to accurately estimate expected returns.Many studies have been conducted to describe the cross-section of expected returns. In some studies, such as Rubinstein (1976), Lucas (1978), and Brindle (1979), fewer variables were considered, and in some, such as Harvey (2016), hundreds of papers and factors with greater predictive power for the expected cross-section of returns were examined. Which of these factors and characteristics have the greatest impact on predicting returns and providing incremental information to the researcher and investor is important to these studies.

According to Cochran (2011), with regard to the introduction and evaluation of new features, different methods should be used to identify effective characteristics in predicting expected returns.In this regard, the purpose of the present study is to determine which of the ordinary Lasso regression methods, the adaptive group Lasso regression, and the ordinary least squares method provide additional information to predict the expected return on the portfolio.

The results of this study indicate that many of the characteristics proposed in the field of expected return on stocks in previous studies have less power and influence on forecasting. Based on the results of the research using ordinary least squares method, 7 characteristicshave the power to predict expected returns. If using the ordinary Lasso method, fewer features were selected than the previous one.In this method, 5 firm characteristics of the stocks were used to predict the expected return on the portfolio. However, we can achieve similar results using the adaptive group Lasso regression method taking into account only 4 characteristics. In other words, using the least effective variables or characteristics, we can predict the expected return.

The results show that investors and financial analysts can use better adaptive group Lasso regression in less time and less costly analysis for investment decisions.Since this research uses annual financial information of companies, it is recommended for future studies to obtain more accurate information and compare it with the results of this research using monthly or seasonal information of companies listed in Tehran Stock Exchange. It is also suggested that other corporate characteristics, as well as macroeconomic characteristics, in particular the effects of inflation and exchange rate changes be examined.

References

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang. (2006).The cross-section of volatility and expected returns. *The Journal of Finance* 61 (1), 259-299.

Avramov, Doron, &Ghordia, Tarun. (2006). "Asset Pricing Models and Financial Market Anomalies". *The Review of Financial Studies*, 19(3). doi:10.1093/rfs/hhj025

Balakrishnan, K., E. Bartov, and L. Faurel. (2010). Post loss/profit announcement drift. *Journal of Accounting and Economics* 50 (1), 20-41.

Bali, T. G., N. Cakici, and R. F. Whitelaw. (2011).Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics* 99 (2), 427 - 446.

Bandyopadhyay, S. P., A. G. Huang, and T. S. Wirjanto. (2010).The accrual volatility anomaly. *Unpublished Manuscript, University of Waterloo*.

Basu, S. (1983). The relationship between earnings' yield, market value and return for NYSE common stocks: Further evidence. *Journal of Financial Economics* 12 (1), 129-156.

Beshkooh and Afshari. (2012). ' the optimal sampling performance of investing in the stock market using a combination analysis of hierarchical analysis (AHP) and analysis of the theory of gray (GRA) ', *Journal of scientific and applied scientific research*, vol. 2, no. 11, pp 11218-11225 [in Persian]

Bhandari, L. C. (1988). "Debt/Equity Ratio and Expected Common Stock Returns: Empirical Evidence. "*Journal of Finance* 43, 507-528.

Bossaerts, P., Hillion, I. (1999). Implementing Statistical Criteria to Select Return Forcasting Models: What Do We Learn?*The Review of Financial Studies*, 12 *(*2 *):* 211-222.

Bryzgalova, S. (2016). Spurious factors in linear asset pricing models. *Unpublished Manuscript, Stanford University.*

Chinco, A., A. D. Clark-Joseph, and M. Ye. (2015). Sparse signals in the cross-section of returns. *Unpublished Manuscript, UIUC.*

Chordia, T., A. Subrahmanyam, and V. R. Anshuman.(2001). Trading activity and expected stock returns. *Journal of Financial Economics* 59 (1), 3-32.

Cochrane, J. H. (2011). Presidential address: Discount rates. *Journal of Finance* 66 (4),1047-1108.

Cooper, M. J., H. Gulen, and M. J. Schill. (2008). Asset growth and the cross-section of stock returns. *The Journal of Finance* 63 (4), 1609-1651.

De Bondt W, R. H Thaler. (1985). "Dose the Stock Market Overreact?" *Journal of Finance*, Vol. 40: 793-805.

Edirisinghe, NCP & X Zhang,(2008). Portfolio selection under DEA-based relative financial strength indicators: case of US industries, *Journal of the operational research society* 59.

Fama, E. F. and K. R. French (1992). The cross-section of expected stock returns. *Journal of Finance* 47 (2), 427-465.

Fama, E. F. and K. R. French. (1996). Multifactor explanations of asset pricing anomalies.*Journal of Finance* 51 (1), 55-84.

Fisher. Black & M. Scholes. (1974)"The Effects of Dividend Yield and Dividend Policy on Common Stock Prices and Returns", *Journal of Financial Economics*

Frazzini, A. and L. H. Pedersen. (2014). Betting against beta. *Journal of Financial Economics* 111 (1), 1-25.

Gandhi, P. and H. Lustig. (2015). Size anomalies in US bank stock returns. *The Journal of Finance* 70 (2), 733-768.

George, T. J. and C.-Y. Hwang. (2004).The 52-week high and momentum investing. *The Journal of Finance* 59 (5), 2145-2176.

Giglio, S. W. and D. Xiu. (2016). Inference on risk premia in the presence of omitted factors. *Unpublished Manuscript, University of Chicago.*

Green, J., J. R. Hand, and F. Zhang. (2016).The characteristics that provide independent information about average US monthly stock returns. *Review of Financial Studies*

Hahn, J. and H. Lee. (2009). Financial constraints, debt capacity, and the cross-section of stock returns. *The Journal of Finance* 64 (2), 891-921.

Harvey, C. R., Y. Liu, and H. Zhu. (2016). ..and the cross-section of expected returns.*Review of Financial Studies* 29 (1), 5-68.

Haugen, R. A. and N. L. Baker. (1996). Commonality in determinants of expected stock returns. *Journal of Financial Economics* 41 (3), 401-439.

Huang, J.-Z and Z. Shi. (2016). Determinants of bond risk premia. *UnpublishedManuscript, Penn State University.*

Janani, M. H.; Ehsanifar, M. & S. Bakhtiarnezhad (2012), "Selection of Portfolio by Using Multi Attributed Decision Making (Tehran Stock Exchange) ", American Journal of Scientific Research, Vol. 44, PP. 15-29

Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *The Journal of Finance* 45 (3), 881-898.

Lakonishok, J., A.Shleifer, &R.W.Vishny. (1994).Contrarian investment, extrapolation, and risk. *The Journal of Finance* 49 (5), 1541-1578.

Lamont, O., Ch. Polk, and S-R. Jesus. (2001). Financial Constraints and Stock Return Volatility. *The review of Financial Studies* 14(2): .529-554

Lewellen, jonathan. (2015). The Cross-section of Expected Stock Returns, Tuck School of Business Working Paper No. 2511246- *Critical Finance Review*, 2015, 4: 1–44

Light, N., D. Maslov, and O. Rytchkov. (2016). Aggregation of information about the cross section of stock returns: A latent variable approach. *Review of*

Financial Studies

Lyandres, E., L. Sun, and L. Zhang. (2008).The new issues puzzle: Testing the investmentbasedexplanation. *Review of Financial Studies* 21 (6), 2825-2855.

Namazi. andRostami. (2005), ' a review of financial ratios and the rate of return on the stock between the TSE in Tehran stock exchange ', *quarterly reviews for accounting and auditing*, no. 44, pp. 127-105. [in Persian]

Novy-Marx, R. (2011). Operating leverage. *Review of Finance* 15 (1), 103-134.

Novy-Marx, R. (2012). Is momentum really momentum? *Journal of Financial Economics* 103 (3), 429-453.

Olsen, Dennis and Charles Mossman. (2003). "Neural Network Forecast of Canadian Stock Returns" *International Journal of Forecasting*, No.19, PP: 453-465.

Ou, J. A. and S. H. Penman (1989). Financial statement analysis and the prediction of stock returns. *Journal of Accounting and Economics* 11 (4), 295-329.

Pontiff, J. and A. Woodgate. (2008). Share issuance and cross-sectional returns. *The Journal of Finance* 63 (2), 921-945.

Palazzo, B. (2012). Cash holdings, risk, and expected returns. *Journal of Financial Economics* 104 (1), 162-185

Pour, zamani and bashiri, Ali. (2013)."Carhart model test to predict the expected return to growth stocks and value breakdown." *Journal of financial engineering and management of securities*, volume 4, no. 16, pp 93-107. [in Persian]

Richardson, S. A., R. G. Sloan, M. T. Soliman, and I. Tuna. (2005). Accrual reliability, earnings persistence and stock prices. *Journal of Accounting and Economics* 39 (3), 437-485.

Rosenberg, B., K. Reid, and R. Lanstein. (1985). Persuasive evidence of market inefficiency. *The Journal of Portfolio Management* 11 (3), 9-16.

Sheari, Saber. (2003): the role of basic accounting information in forecasting stock returns, Faculty of accounting and management, *allameTabatabaie University, PhD thesis* [in Persian]

Shah Nazari, m. r. (2004). explaining alternative systematic risks criteria, *Master thesis, Islamic Azad University, science research.*

Soliman, M. T. (2008). The use of DuPont analysis by market participants. *The Accounting Review* 83 (3), 823-853.

Sloan, R. (1996). Do stock prices fully react information in accruals and cash owes about future earnings? *Accounting Review* 71 (3), 289-315.

Uddin, R.; Rahman, Z. & R. Hossain. (2013). Determinants of Stock Prices in Financial Sector Companies in Bangladesh- a Study on Dhaka Stock Exchange (DSE) , *Interdisciplinary Journal of Contemporary Research In Business*, Vol. 5, No. 3, PP.

Zamazi, r. (2010). The relationship between financial ratios based on the cash flow and stock returns Tehran Stock Exchange. *Master thesis, AllamehTabatabai University, Tehran*. 471-480.

Bibliographic information of this paper for citing:

Mortazavi, Raheleh ossadat; Vakilifard, Hamid Reza; Talebnia, Ghodratallah & Jafari, Mahboobeh (2018). Comparison of linear regression models Ordinary Lasso, Adaptive Group Lasso and Ordinary Least Squares models in selecting effective characteristics to predict the expected return. *Iranian Journal of Finance*, 2(3), 49-69.