Iranian Journal of Management Studies (IJMS) *http://ijms.ut.ac.ir/*
Vol. 11, No. 4, Autumn 2018 *Print ISSN: 2008-7055*
pp. 693-714 *Online ISSN: 2345-3745*
DOI: 10.22059/IJMS.2018.242990.672848

Investigating Predictability of Different "Forms of Return" in Tehran Stock Exchange: Some Rolling Regressions-based Evidence

Azam Mohtadi[1&2*], Rezvan Hejazi[1], Sayed Ali Hosseini[1], Mansoor Momeny[3]
1. Accounting, Faculty of Social Sciences and Economics, AL Zahra University, Tehran, Iran
2. Accounting, Department of Management, Economic and Accounting, Payame Noor University, Tehran, Iran
3. Industrial Management/ Faculty of Management, University of Tehran, Tehran, Iran

(Received: December 2, 2017 ؛ Revised: August 23, 2018; Accepted: September 12, 2018)

Abstract
This paper has provided "out of sample" evidence of stock returns predictability in Tehran Stock Exchange. 68 qualified companies over the period from 2002 to 2015 were selected and for five different "forms of returns", five superior predictive models have been designed by applying "General to specific" approach of modeling technique. Then "out of sample" analysis, based on rolling regressions, has been used to test the validation of the designed models. The result showed that all designed models have sufficient "out of sample" validity and the aggregate returns have a higher predictability level.

Keywords
Returns, out of sample, Rolling Regressions, General to Specific

* **Corresponding Author, Email:** a_Mohtadi@yahoo.com

Introduction

Researchers for various reasons, such as finance-behavioral theories or absence of full market efficiency, have considered stock returns to be predictable, and for many years (for example, Dow, 1920) have attempted to identify the nature and elements of its formation. This article is also on this topic and tries to investigate the predictability of the stock returns in Iranian texture.

So far several models, such as Ou & Penman (1989), Fama & French (1993), Zhang (2002), Fama & French (2015), have been proposed, in order to predict capital market company's returns, but all of these models originate in other countries and there is evidence that the returns related models are not accurately applicable in different countries (Griffin, 1997; Thomsen, 2010). Furthermore, the elements of the variables that form the models change over time (Sheari, 2004; Abbaszadeh & Golestani, 2010; Paye & Timmermann, 2006) and people with different investment horizons participate in the market. Each of them focuses on aspects of returns. For example, while a short-term investor is interested in information about next year returns, mid-term and long-term investors are also likely to pay attention to the aggregate returns. Therefore, in addition to country-specific predictive models, prediction of each type (form) of returns, by meeting the information needs of a particular range of investors, will be beneficial.

These reasons and paying attention to the conditions of Iran, such as double-digit inflation, unusual downturns and booms in alternative markets and special political conditions raised this possibility that variables related to the returns in Iran are different from other countries and motivated us to design Iranian specific stock returns models. Specifically, in this research it has been attempted to investigate the predictability of some forms of returns, including three forms of aggregate returns (as the representative of the midterm returns), next year returns (as the representative of the short-term turn) and one year delayed returns (as a factor affecting the inefficiency of the capital market in Iran).

Usual regression models that are based on data mining techniques tend to look-ahead bias (Zhong & Enke, 2017). Rolling regressions is one of the available methods to test the "application of the models in practice", and is rooted in the concept of "out of sample" analysis

(Thomsen, 2010). In the process of rolling regressions, it is assumed that the person is in the real world and at any time (year), using all the available information, designs his/her model and predicts the future; therefore, using this method will reduce the data mining problems. We design, compare, test, and rank models with the "out of sample" analysis based on rolling regressions technique to prepare a comprehensive analysis of stock returns predictability in Iranian texture. Comprehensiveness of the present research about the analysis of several forms of returns and the use of rolling regressions in models validation are considered as our contribution to the literature.

Our paper now proceeds as follows. We discuss the theoretical foundations of the work in the following section. In next three sections we first explain the data and analysis methods then design validate and rank our models. Finally, a summary of results and conclusions are presented.

Theoretical Foundations

Arbitrage Pricing Theory

Arbitrage pricing theory (APT) is a well-known method of estimating the price of an asset. The theory assumes an asset's return is dependent on various macroeconomic, market and security-specific factors. The general idea behind APT is that two things can explain the expected return on a financial asset: 1) macroeconomic/security-specific influences and 2) the asset's sensitivity to those influences. This relationship takes the form of the linear regression formula. In general, using the assumptions of (APT) provides sufficient backing to predict stock returns with a set of relevant variables. Some researchers predict stock returns based on this theory. (Valizadeh, (2014), Namazi & Mohammad Tabar Kasegari, (2007) and others)

Literature review

Despite numerous researches, systematic efforts to design the forecast models of returns in Iran are limited. Valizadeh (2014) in his doctoral thesis has presented a model for predicting next year stock returns. She used confirmatory factor analysis to confirm the initial structure of her data and presented her model with the "specific to general" regression technique. In her model, there was a combination of economic, corporate and market variables. She claimed to have presented a model for explaining next period stock returns.

Ali Mohammadi et al. (2015) presented a model for forecasting current and next year returns using financial ratios. They used the "decision tree" approach to design their model and stated that "decision tree" approach is applicable in model designing.

Recently, Setayesh & Kazemnejad (2016) in their research initially acted to form models for forecasting the returns and then compared the models with each other. They initially identified 52 variables and then ranked them by correlation and relief techniques, and finally, they used seven variables that had the most relationship to returns as variables in their models (Judicial Approach).Their results showed that the use of ensemble models predicts returns better than simple models.

Welch & Goyal (2008) after a massive review of the US capital market found that "out of sample" evidence (rolling regressions) does not confirm the predictability of returns. Campbell & Thompson (2007) applied some limitations in the calculations of these researchers and rejected their result by achieving weak but significant evidence.

Thomsen (2010), Zhou & Faff (2017), and others provided "out of sample" evidence and confirmed "out of sample" predictability of returns.

A review of domestic researches on the modeling of stock returns highlights two points: First, domestic investigations conducted to date often rely on internationally known models such as Fama & French (1993), and the efforts on modeling the returns based on country specific characteristics are limited. Researchers often analyze one or two forms of returns and a comprehensive review of different forms of returns with a vast data set has not been considered.

So in this study, the following question and related hypotheses are tested:

1. By observing the principles of General to specific procedure in modeling, what are the best models (combination of variables) for predicting each form of stock returns in Iran?

To answer this question, 120 sub hypotheses are tested (24 variables in 5 models). One example of these hypotheses is as follows:

H0: There is a significant relationship between "assets growth" and one period future returns.

The second highlighted point is about researches' techniques. Despite much evidence regarding the application of the rolling regressions techniques to validate regression models in international studies, this subject has not been considered in domestic researches. Domestic researchers often rely on in-sample testing of their models, and there is no strong experimental evidence in models validating based on "out of sample" evidence. So in this study, a second question and group of hypotheses are tested.

2. Do designed models have "out of sample" validity?

To answer the second question, we apply "out of sample" approach and test the hypothesis below for each kind of returns :

H0: Based on the pattern of rolling regressions in the "out of sample" validation technique, for "each form of the returns", the errors of the "historical mean model" are less than or equal to the errors of designed model.

The term "each form of returns" means five forms of returns that are investigated in this study. (See last 5 rows of table 1).In the end, as the complimentary information model is ranked based on Theil (1966) measures of accuracy and quality.

Methodology

Sample and Data

Listed and active companies in Tehran Stock Exchange have formed the statistical population of the research. In this research, sampling has not been performed, and only the companies that have following situations have been omitted from the studied population:

1. Investment companies and banks were omitted from the studied population.
2. The companies that their end of the fiscal year does not fit into the real year, as well as those companies that changed the fiscal year during the years studied, were omitted from the surveys.
3. Unprofitable companies have been omitted from the studied population.
4. Companies that due to reasons such as the lack of trade throughout the year, temporary closeness, withdrawal from the stock exchange, and other reasons had no suitable data for analysis were omitted from the investigations.

Finally, by applying the above conditions, the studied companies were limited to 68 companies. In this research, the analyses have been conducted annually, and the beginning of the period under consideration is returned to one year after the required period for the establishment of accounting standards of Iran (2001) and specifically, this period includes years from 2002 to 2015 (14 years). Required data were extracted from professional software (Novin Software) and other credible sources of information, such as the Stock Exchange library and official Cite on Iranian national bank. Excel 2010, Eviews 9 and Stata 14 software were used to analyze the data.

Research Variables

Research variables were identified by a deep study of the literature both inside and outside the country. These variables include 47 variables, which are based on the three fields of company, market, and economy. Definitions and descriptive statistics for all variables (47 independent and five dependent variables) have been presented in Table 1. In this table, large numbers are presented in thousands of billions of Rials.

Due to a large number of research variables, the use of the symbol will be confusing; so in this table, for each variable, a code has been introduced which variable will be identified with throughout the paper. To limit the effect of outliers, the variables have been winsorized at the 5% level of upper and lower limits.

Table 1. Descriptive information and operational definitions of research variables Large numbers are presented in thousands of billions of Rials.

code	Variable Symbol	Operational definition	Mean	Std	Max	Min
1	Market Value	Number of shares × Share value	1.03	1.60	6.95	0.06
2	Accrual Items	(Total assets - Cash &investments) - (Total debts- Current debts)	0.84	1.10	4.73	0.08
3	Return on Assets	Net income /Total assets	0.16	0.11	0.40	0.02
4	Return on Working Capital	Net income/Working capital	0.47	1.72	4.10	-3.57

Table 1. Descriptive information and operational definitions of research variables Large numbers are presented in thousands of billions of Rials (Continiouse)

code	Variable Symbol	Operational definition	Mean	Std	Max	Min
5	Return on Equity	Net income/Stock holders' equity	0.41	0.22	0.85	0.07
6	Beta	Systematic risk measure ($cov_{i,m}/var_m$)	0.39	0.93	2.46	-1.03
7	ΔCash	Cash and its equivalents$_t$ - Cash and its equivalents$_{t-1}$	0.06	0.025	0.08	-0.05
8	Δ Current Value	Market value$_t$ - Market value$_{t-1}$	0.18	0.61	2.20	-0.94
9	Total Assets	Total assets	1.03	1.45	6.35	0.09
10	%Income	Net income/Sales	0.22	0.17	0.67	0.02
11	%Operational Profit	Operational profit/Sales	0.26	0.15	0.61	0.05
12	%ΔConsumer Price Index(CPI)	$(CPI_t – CPI_{t-1})/(CPI_{t-1})$	0.19	0.08	0.35	0.10
13	%Δ Gold Price	(Average of Gold Prices$_t$- Average of Gold Prices$_{t-1}$)/ (Average of Gold Prices$_{t-1}$)	0.30	0.23	0.80	-0.09
14	%Δ Average of Oil Prices	(Average of Oil Prices$_t$ - Average of Oil Prices$_{t-1}$)/ (Average of Oil Prices $_{t-1}$)	0.15	0.17	0.38	-0.56
15	%Payout Ratio	Dividend per share/Earnings per share	0.69	0.26	1.00	0.10
16	%Institutional Investors	Percentage of shares owned by banks, financial institutions, Insurance companies, and all real and legal persons, with more than 5% of shares.	0.77	0.17	0.970	0.26
17	% Un-Institutional Investors	1- % Institutional investors	0.23	0.17	0.73	0.03
18	Non-Governmental Public Ownership	1-Percentage of shares owned by government agencies	0.11	0.31	1.00	0.00
19	Assets Growth	$(TA_t – TA_{t-1})/(TA_{t-1})$	0.21	0.19	0.62	-0.07
20	Sales Growth	$(Sales_t – sales_{t-1})/(Sales_{t-1})$	0.24	0.27	0.92	-0.22
21	Net Working Capital	Current assets – Current liabilities	0.13	0.25	1.00	-0.33

Table 1. Descriptive information and operational definitions of research variables Large numbers are presented in thousands of billions of Rials (Continiouse)

code	Variable Symbol	Operational definition	Mean	Std	Max	Min
22	Operational Profit(Loss)	Sales – The cost of goods sold – Other operational expenses	0.19	0.27	1.18	0.009
23	Income After Tax	Net income after tax	0.16	0.27	1.17	0.006
24	Income Before Tax	Net income before tax	0.19	0.30	1.32	0.007
25	% Net to Gross Income	Net income/gross income	0.62	0.23	1.35	0.13
26	Dividend per Share(Rials)	Dividend/Number of outstanding shares	860	845	3018	30
27	Earnings Per Share(Rials)	Net income/Number of outstanding shares	1118	933	3433	119
28	$E/P_{(t-1)}$ Ratio	Earnings per share $_t$ /Price of a share $_{t-1}$	0.20	0.12	0.46	0.031
29	Group Membership	Dummy variable which takes 1 when a company is member of a commercial group and 0 otherwise	0.74	0.44	1.00	000
30	Price to Sales Ratio	Share Price/Sales per capita	1.73	1.48	5.69	0.28
31	Total Assets Turnover	Daily Sales/Average total assets	0.81	0.30	1.45	0.23
32	Family Membership	Dummy variable which takes 1 when a special family has more than 20 % of a company's shares and 0 otherwise	0.15	0.36	1.00	000
33	Momentum	Dummy variable which takes 1 when net income of a company is more than last year and 0 otherwise	0.68	0.47	1.00	000
34	Interest Rate	Interest expenses/Total debts	0.05	0.03	0.13	000
35	Effective Tax Rate	Tax expenses/Total sales	0.03	0.03	0.10	000
36	Quick Ratio	Current assets (except inventories and investments)/Current liabilities	0.79	0.34	1.67	0.26

Table 1. Descriptive information and operational definitions of research variables Large numbers are presented in thousands of billions of Rials (Continiouse)

code	Variable Symbol	Operational definition	Mean	Std	Max	Min
37	Book to Market Ratio	Book value of firm /Market value of firm	0.57	0.38	1.4	0.13
38	Debt Ratio	Total debts/Total assets	0.62	0.14	0.83	0.31
39	Debt to Equity Ratio	Total debts/Stock holders' equity	2.01	1.20	4.97	0.45
40	Current Ratio	Current assets/Current liabilities	1.29	0.46	2.47	0.59
41	Current Assets Ratio	Current assets/Fix assets	0.66	0.17	0.89	0.28
42	Dividend-Price Ratio	Dividend per Share/Market value per share	0.13	0.07	0.28	0.01
43	Price-Earning Ratio	Market value per share/Earnings per share	7	3.88	19.6	2.7
44	Earnings-Price Ratio	Earnings per Share/Market Value per Share	0.1	0.08	0.37	0.05
45	Liquidity	Days with at least one deal/total trading days of the market.	0.04	0.02	0.08	0.01
46	Cash Flows	Operational cash flows	0.13	0.21	0.91	-0.02
47	%Δ Volume of Money (VOM)	(The average of $Vom_{(t)}$ - The average of $Vom_{(t-1)}$)/ The average of $Vom_{(t-1)}$	0.19	0.34	0.39	-0.9
R_{t+1}	Returns of the Next Year	The returns on common stock during the next year (period t+1)	0.36	0.63	2.13	-0.32
R_{t+2}	One Year Delayed Returns	The returns on common stock during the second next year (period t+2)	0.37	0.65	1.95	-0.48
R_{com2}	Aggregate Returns 1	The sum of the current and next year returns	0.80	0.93	3	-.36
R_{com3}	Aggregate Returns 2	The sum of the current and two next years' returns	1.27	1.09	3.57	-.025
R_{com2R}	Aggregate Returns 3	The sum of the two next years' returns	0.79	0.93	2.97	-0.36

Data analysis method

The method used in this research is the descriptive-correlational type. Designing approaches to regression models and rolling regressions approaches has been used in the panel data context.

The method of conducting the research was as follows. First, the variables that have theoretical relation with returns were identified with a thorough study of the literature. Subsequently, using the "Principle Components Analysis" technique, the volume of initial data set was reduced to make possible the design of the models with “general to specific" (GS) approach. Then, the most suitable predictive model was designed using the mentioned approach, for each form of returns and finally, extracted models were validated using an "out of sample" approach based on rolling regressions and ranked based on the forecast accuracy and quality.

Designing a model in econometrics

In econometrics, two "specific to general" (SG) and "general to specific" (GS) methods have been proposed for designing a model and selecting among many variables (Aflatuni, 2014). The basis for doing the work in the "SG" method is to arrive from a basic and simple model to the ultimate and powerful model. Hendry and Richard (1982) in addition to presenting the criticisms to "SG" method presented the "GS" approach (Armstrong, 2001). In this approach, the researcher starts the work with a comprehensive model that consists of all the variables that affect the studied subject. The mentioned model is fitted and proper validation tests are carried out to determine its reliability. Then, it is tried to simplify the model and provide a final model by omitting variables that are not statistically significant.

Principle component analysis

So far, the literature on stock returns has introduced about 50 variables related to returns (Setayesh & Kazemnejad, 2016). Using all of these variables in the primary regression of "GS approach" is not possible for two reasons; first, co-linearity will exist between some of the variables and second, excessive reduction of the degree of freedom might be prevalent. Therefore, it is necessary to use valid data-reduction methods to reduce the initial number of variables to an acceptable level. Sorzano et al. (2014) stated that among the techniques for reducing data dimensions, the "Principal component analysis" (PCA) method and its various versions are simpler and more comprehensible than other methods. Hargreaves and Mani (2015), Wang & Choi (2013) successfully used the (PCA) method for data reduction of variables affecting stock returns. In the "PCA" method, by applying the concepts of "Eigen values" and "Eigen vectors", several correlated variables are converted to one (or several) un-correlated component(s) and thus, in the final analysis, the volume of data decreases. In this study, we used "PCA" to reduce our variables.

"Out of sample" analysis based on rolling regressions

Various researchers such as Campbell and Thompson (2007), Thomsen (2010), Bahrami et al. (2016) and others, have used "out of sample" methods to perform the generalizability of models with more reliability.

Rolling regressions technique (RR) is rooted in the concept of "out of sample" analysis. The method of doing the work is that at first, the study period T is divided into two categories, including the "First Period to Period F" and "Period F to Period T" (Equal or Non-equal). Typically, the "first period to period F" is called in-sample period, and the second period is called the "out of sample" period. Then, the returns and so the errors of each year of ""out of sample"" period are calculated and finally, by comparing the errors of the designed model with a benchmark model, predictability of the designed model is concluded.

The estimate of the returns and errors for each year of the "out of sample" period is as follows: first, with the data of in sample-period (first period to period F), the first regression equation is estimated and the returns of the F + 1 period is forecasted. Then, by decreasing the real returns from the forecasted returns, the errors of period F + 1 is calculated. Then the estimation window is stepped forward one period and with data related to the “first period until period F + 1" the second regression equation is estimated and used to forecast the returns and errors of the F + 2 period with the same pattern of the previous period; this approach continues until the last examined period (period T).

The model that has been used in this research as the benchmark model is "historical average returns model". This model usually used as the benchmark model in returns related research. Welch and Goyal (2008), Campbell and Thompson (2007) and Thomsen (2010) and many others used this model as their benchmark model . The main concept of the "out of sample" approach has been depicted in Table 1. In this fig, the period 2002 to 2005 is the in-sample period and the periods 2006 to 2015 are considered as the "out of sample" periods.

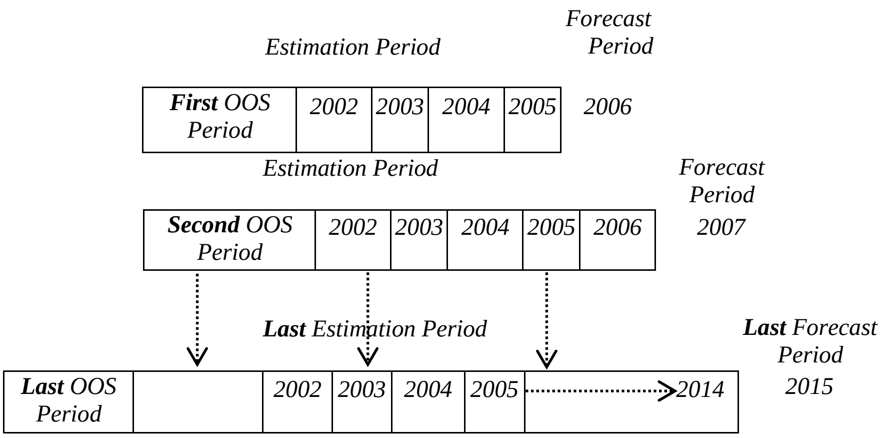

Fig. 1.The concept of rolling regretions

"Out of sample" tests

"Out of sample" R squared

The determination coefficient (R2) determines the statistical significance of the models based on the "out of sample" period. This coefficient is calculated from equation 1 (Thomsen, 2010; Welch & Goyal, 2008):

$$R_{OS}^{2} = 1 - \frac{\sum_{i=1}^{q}\left(r_{F+i} - \hat{r}_{u,F+i}\right)^{2}}{\sum_{i=1}^{q}\left(r_{F+i} - \hat{r}_{r,F+i}\right)^{2}} \qquad (1)$$

In this model, r^_(u,F+i) is the forecasted returns based on the designed model and r^_(r,F+i) is the forecasted returns by the benchmark model. The decision rule is that if the calculated coefficient is greater than zero, it means that the designed model has had a forecasted mean squared error less than the historical model and can be used as a forecasted model.

McCracken (2007)

The test of McCracken (2007) is a famous test at the time of comparing nested models. The confirmation of the null hypothesis of McCracken test (2007) means that the designed model has no suitable forecast ability. In this research, this test has been used for comparison of the designed (unrestricted) and benchmark (restricted) models.

Theil measures (1966)

Also, in this research, two Theil accuracy coefficient (TIC) and Theil quality coefficient (TIC-UII) have been used to rank the models. According to the formula, the value of these coefficients had been between zero and one, and as much the number of these criteria to be closer to zero, the model has better forecast accuracy and quality (Hartmann et al. 2014). In Theil's models (equations (2) and (3)), y_it is the actual return on the company (i) in the year (t) and (y_it)ˆis the predicted return of the same company and the same year and n is the number of observations.

$$\mathrm{TIC-UI} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_{it} - \widehat{y_{it}})^2}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_{it})^2} + \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\widehat{y_{it}})^2}} \qquad (2)$$

$$\mathrm{TIC-UII} = \frac{\sqrt{\sum_{i=1}^{n}(\widehat{y_{it}} - y_{it})^2}}{\sqrt{\sum_{i=1}^{n}(y_{it})^2}} \qquad (3)$$

Besides these criteria, a various set of statistical criteria has been used to achieve the reliable results. These methods and criteria have been presented in summary in table 2.

Table 2. Summary of Tests Used in this Research

Goal(s)	General Technique(s)	specific Test(s)
Data Reduction of Variables	Principles component analysis	Bartlett's test of Sphericity Kaiser-Meyer-Olkin Measure of Sampling adequacy or (MSA measure) Eigen values number
Designing Models	General to specific approach in Panel data context	Limier(chow)test Breusch- test hausman Test
Robustness Test & Basic Validation (Goodness of Fit & Coefficients Overall Validation)	Suitable test	Unit root>>>>> Levin, Lin & Chu Test(LLC) Collinearity>>>>>Variance inflation Factor (VIF) Homoscedasticity>>>>>> Breusch- Pagan-Godfrey Test Serial correlation(SC) >>>> Breusch- Godfrey SC LM Test F test, Adj R^2
"out of sample" Validation	Rolling regressions in panel data context	OOS R^2 McCracken 2007 or MSE-F Test
Models Comparison	Rolling regressions in panel data context	UI Test of Theil(1966) UII Test of Theil(1966)

Research results

Determining the final set of research variables with PCA

The final result of applying the "PCA" method is the conversion of 27 variables (more than 50%) to five independent components. The results have been obtained after applying all necessary tests which have been mentioned in the first panel of table 2. Final estimation equation of each component has been presented in Table 3. Simply, these equations are composed of highly correlated variables (based on PCA concepts) and are used to estimate the number of each component. The estimated components then have been replaced by the constituent variables in the analysis.

Table 3. final models of components

Optional Component Name	models
PPS	$\mathrm{Pps} = 0.25\boxed{28} + 0.28\boxed{42} - 0.30\boxed{43} + 0.32\boxed{44}$
CN1	$\mathrm{CN}_1 = 0.14\boxed{1} + 0.13\boxed{2} + 0.13\boxed{9} + 0.10\boxed{21} + 0.14\boxed{22} + 0.14\boxed{23} + 0.14\boxed{24} + 0.03\boxed{36} - 0.03\boxed{38} - 0.03\boxed{39} + 0.03\boxed{40} + 0.12\boxed{46}$
CN2	$\mathrm{CN}_2 = 0.03\boxed{1} + 0.09\boxed{2} + 0.09\boxed{9} - 0.11\boxed{21} + 0.04\boxed{22} + 0.02\boxed{23} + 0.02\boxed{24} - 0.25\boxed{36} + 0.27\boxed{38} + 0.25\boxed{39} - 0.28\boxed{40} + 0.05\boxed{46}$
CPR1	$\mathrm{CPR}_1 = 0.14\boxed{3} + 0.13\boxed{5} + 0.13\boxed{10} + 0.13\boxed{11} + 0.07\boxed{15} + 0.10\boxed{25} + 0.13\boxed{26} + 0.13\boxed{27} + 0.11\boxed{30} + 0.10\boxed{35} - 0.09\boxed{37}$
CPR2	$\mathrm{CPR}_2 = -0.003\boxed{3} - 0.21\boxed{5} + 0.38\boxed{10} + 0.23\boxed{11} - 0.15\boxed{15} + 0.36\boxed{25} - 0.27\boxed{26} - 0.26\boxed{27} + 0.33\boxed{30} - 0.13\boxed{35} + 0.31\boxed{37}$

The numbers inside the square are variable codes which have been presented in Table 1

The components are composed with 19 other variables, (collectively 24 variables) the final set of variables used in the design of forecast models.

Designing forecast models with a general to specific approach

In this phase, first, the stationarity of variables was tested by Levin, Lin & Chu measure and then in total, by fitting and modifying the fifteen regression equations, the final model of all forms of returns (five forms) has been designed. To summarize, in each case only the final model and its initial validation measures (Adj R2 and F Test) have been presented and the equation of the previous levels has not been provided. (Table 4)

The tests used to examine the regression hypotheses are the cases that have been mentioned in Table 2. If necessary, the proposed method of Aflatuni (2016, 316) has been used to correct the regression problems.

Table 4. Final models designed with the general to specific approach

Dependent variable	Final model
Next Year Returns	$r_{t+1} = -46.6 + 7.4(8) + 124.9(13) - 118.9(14) + 31.6(31) + 13.7(33) + 127.9(34) + 78.3(47) + 4(cpr1) - 5.2(cn1) + 6.5(cn2) + 15(pps) + 11.2(cpr2) + \varepsilon$ *Regressions type: Panel data fixed effect model --- (Adj R^2: 46%, F test P-Value:0.0)*
Current and Next Year Aggregate Returns	$r_{com2} = -131.3 + 3.11(8) + 387.9(12) - 108.3(14) + 47.4\,(31) + 18(33) + 337.5(47) + 26.2(pps) + 11.6(cpr2) - 12.50(cn1) + \varepsilon$ *Regressions type: OLS* --- (Adj R^2: 55%, F test P-Value:0.0)
One Year Delayed Returns	$r_{t+2} = 58.2 - 5.21(6) + 124.6(12) + 77.6(13) + 124.4(14) - 24(19) + 180.6(34) - 321.6(47) + 5.33(pps) + \varepsilon$ *Regressions type: OLS* --- (Adj R^2: 36%, F test P-Value:0.0)
Current and Two Next Year Aggregate Returns	$r_{com3} = 2.3 + 47.7(8) + 343.9(12) + 141.2(13) - 70(14) + 44.5(20) + 239.4(34) + 470(45) - 103.7(47) - 14.5(cn1) + 12.6(pps) + \varepsilon$ *Regressions type: Panel data fixed effect model* --- (Adj R^2: 44%, F test P-Value:0.0)
Two Next Year Aggregate Returns	$r_{com2R} = 47 + 167.5(12) + 212(13) - 13.11(33) + 420.2(34) - 281.3(47) + 18(pps) + \varepsilon$ *Regressions type: Panel data fixed effect model* --- (Adj R^2: 54%, F test P-Value:0.0)

The numbers inside the Parentheses are variable codes which have been described previously in Table 1

Validation of models with rolling regressions technique

Other researchers have used 25 to 50 percent of the data for the initial estimate (Hartmann et al., 2014; Thompson, 2010; Rapach & Wohar, 2006). By considering the nature of our data, the initial estimation window (W) was determined equal to four periods (272 companies - years, 33% of the total data). To do rolling regressions, the expansion period has been set to 1(step=1). Nine forecast equations were formed for each form of returns and in total the returns and errors of the models were predicted and calculated using 45 regression equations. The table 5 shows "out of sample" determination coefficient (R2) and McCracken test of 2007 in the field of comparing the models designed with the historical returns mean. The star sign beside each statistic means the significant of the provided statistics (rejecting the null hypothesis).

In the case of all forms of returns, the null-hypothesis of McCracken test (2007) is rejected; that indicates the superiority of designed models on historical mean based models. The "out of sample" determination coefficient also confirms this matter and indicates that the forecasting power of these models is much higher than that of their own benchmark model. The result of these two tests is that models have more informational value than making the decision based on the mean of returns.

In Table 6, the ranking result of the models has been presented with two criteria of Theil(1966) forecast accuracy and quality. Table 6 shows the "aggregate returns of three periods" which has higher predictability than other forecast criteria. Theil (1966) criteria show relatively good forecast accuracy and moderate forecast quality for models.

Table 5. OOS R2 and McCracken (2007) test

Model	MSE_F	OR^2
R_{com2}	5.60*	0.62
R_{com2R}	5.15*	0.57
R_{com3}	5.12*	0.56
R_{t+1}	5.10*	0.56
R_{t+2}	4.75*	0.53

Table 6. Theil (1966) test results

Model	TIC-UI	TIC-UII
R_{com3}	0.22	0.41
R_{com2}	0.23	0.44
R_{com2R}	0.26	0.48
R_{t+1}	0.32	0.56
R_{t+2}	0.32	0.57

Summary and Conclusion

This research has focused on a comprehensive analysis of the predictability of returns forms. The results of the research show that all the different forms of returns are somewhat predictable (a positive answer to the first research question). In total, 17 variables (see table 4), formed forecast models of five forms of returns, and the significant relationship of other variables with returns were rejected. The forecast through all forms of returns will be more effective than the use of the moving average and will produce results with lower error (confirmation of the hypotheses related to second question), so that about most successful forecast model, the designed model error includes only about 38 % of the error in the averaging method .This criterion in the worst case reaches around 57%.(see table 5) However, in the accuracy and forecast quality dimension(the third question of the research), models have acted slightly weaker. In the dimension of forecast quality (UII criteria-table 6), at best form (aggregate return with three-periods), the mean of forecast quality is about 60% and in the worst form (one year delayed returns) is about 40%, which are relatively unsuitable and show the medium(weak) quality of forecasts. From the accuracy point (UI criteria-table 6), the numerical index varies from 68% to 78%, which is relatively acceptable.

Other results that have been obtained from the implementation of this research are discussed below:

1. Almost in all conducted analyses and all metrics computed in this paper (tables 4, 5 and 6), the aggregate returns are superior to single-period returns. The reason for this matter is probably that the aggregate returns, with the sum of the returns of several years, will modify the effect of incidental fluctuations that cause to distort results of forecasts and make forecasts more possible. Given this result, it is suggested that in terms of stability and better predictability of investment in the midterm and long term, sufficient awareness is needed to replace the horizons of longer-term investment with short-term horizons and reduce the effects of stock fluctuations which are nearly common in Iran.

2. Another important result of the present research is the emphasis on the significant role of economic variables of oil, gold, inflation and economy liquidity (volume of money) on the returns of companies. These variables had a significant effect on almost all studied models. Valizadeh (2014) emphasized the importance of these variables by including them in his final model. Also, Sajjadi et al. (2010) and Mashayekhi et al. (2010) and others pointed to the significant relationship between some

macroeconomic variables and stock cash returns. In contrast, Namazi & Mohammad Tabar Kasegari (2007) result stated that there is not a significant relationship between economic indexes and the stock returns. However, most of the researches have supported the effect of economic variables on stock index. The results of this research show that the lack of attention to these markets as complementary and substitute markets can encounter regression results with a serious challenge. Therefore, it is recommended that these economic variables be included in the return models so that the research not to be encountered with the omitted variable and more reliable results to be presented. In the international dimension, also several studies such as Rangvid (2006); Cooper & Priestley (2009); Thomsen (2010); Westerlund et al. (2015) have supported the significant effect of economic variables on the stock returns.

3. The significance of some company-based variables such as "assets growth" or "assets turnover" shows that financial reporting has informational content and can be considered at least as one source of information to the market. Though, the ability of the models in predicting all forms of returns, and specially the significance of momentum variable (code 31) in some models challenges the notions of stock market efficiency. In this matter, additional researches are needed and suggested.

4. The stock data reduction in domestic studies is principally based on judicial approaches and little attention has been paid to valid data mining methods, such as "factor analysis" or "principal component analysis". The result of this research showed that the "principal component analysis" technique could be used in data reduction of effective variables on stock returns in Iran's capital market.

Limitations and Areas for Future Research

This research has been carried out in the absence of any significant limitations and the results can be expanded in different aspects such as limiting the components identified by the factor analysis approach, comparing different methods of designing the regression models in econometrics and applying valid data reduction approaches such as "principal component analysis" or "factor analysis" to the other important variables of the capital market such as dividend growth and cost of capital.

References

Abbaszadeh, M. R., & AtashiGolestani, H. (2010). Accounting variables and stock return's forecast before and after the Iranian accounting standards. *Knowledge and Development*, 18(33), 1-28. (In Persian)

Aflatuni, A. (2014). *Statistical analysis in financial management and accounting researches by Eviews*. Tehran: Terme Publication. (In Persian)

Aflatuni, A. (2016). *Statistical analysis in accounting and finance using STATA*. Tehran: Terme Publication. (In Persian)

Alimohammadi, A., Abbasi e Mehr, M. H., & Javahery, A.(2015). Prediction of stock return using financial ratios: A decision tree approach. *Journal of Financial Management Strategy*, 3(4), 124-146. (In Persian)

Armstrong, J. S. (Ed.). (2001). *Principles of forecasting: a handbook for researchers and practitioners* (Vol. 30). Springer Science & Business Media.

Bahrami, A., Shamsuddin, A., & Uylangco, K. (2016). Out-of-sample stock returns predictability in emerging markets. *Accounting & Finance*.

Campbell, J. Y., & Thompson, S. B. (2007). Predicting excess stock returns "out of sample": Can anything beat the historical average? *The Review of Financial Studies*, *21*(4), 1509-1531.

Cooper, I., & Priestley, R. (2009). Time-varying risk premiums and the output gap. *Review of Financial Studies*, 22(7), 2801-2833.

Dow, C. H. (1920). *Scientific Stock Speculation*. Magazine of Wall Street.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, *33*(1), 3-56.

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116 (1), 1:22.

Griffin, J. M. (1997). *Determinants of the cross-section of expected stock returns in Japan*. (Doctoral dissertation, The Ohio State University).

Hargreaves, C. A., & Mani, C. K. (2015). The Selection of winning stocks using principal component analysis. *American Journal of Marketing Research*, *1*(3), 183-188.

Hartmann-Wendels, T., Miller, P., & Töws, E. (2014). Loss given default for leasing: Parametric and nonparametric estimations. *Journal of Banking & Finance*, *40*, 364-375.

Hendry, D. F., & Richard, J. F. (1982). On the formulation of empirical models in dynamic econometrics. *Journal of Econometrics*, 20(1), 3-33.

Mashayekhi, B., Tahriri, A., Ganji, H., & Askari, M. R. (2010). The impact of macroeconomic variables on the relation between fundamental variables derived from financial statements and stock returns. *Quarterly Journal of Securities Exchange*, 3(12), 109-126. (In Persian)

McCracken, M. W. (2007). Asymptotics for "out of sample" tests of Granger causality. *Journal of Econometrics*, 140(2), 719-752.

Namazi, M., & Kasgary, H. (2007). Using a multi-factor model to explain stock returns of companies accepted in Tehran Stock Exchange. *Journal of Accounting Advances*, 26(1), 157-180. (In Persian)

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

Paye, Bradley S., and Allan Timmermann, (2006). Instability of return prediction models. *Journal of Empirical Finance*, 13, 274-316.

Rangvid, J. (2006). Output and expected returns. *Journal of Financial Economics*, 81(3), 595-624.

Rapach, D. E., Strauss, J. K., & Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *The Review of Financial Studies*. 23(2), 821-862.

Rapach, D. E., & Wohar, M. E. (2006). In-sample vs. out-of-sample tests of stock returns predictability in the context of data mining. *Journal of Empirical Finance*, *13*(2), 231-247.

Sajadi, H., Frazmand, H., & Alisufi, H. (2010). Investigating the relationship between macroeconomic variables and exchange price index in Tehran stock exchange. *Accounting Research*, 10(2), 123-150. (In Persian)

Setayesh, M. H., & Kazemnejad, M. (2015). Investigating usefulness of ensemble regression and feature selection in predicting stock returns of companies listed on Tehran stock exchange. *Journal of Financial Accounting and Auditing Research*, 8(32), 1-28. (In Persian)

Sheri, S. (2004). *The Role of Fundamental Accounting Information in Predicting Stock Returns*. (Doctoral dissertation), Allameh Tabatabaei University,Iran. Retrieved from: library.atu.ac.ir. (In Persian)

Sorzano, C. O. S., Vargas, J., & Montano, A. P. (2014). A survey of dimensionality reduction techniques. arXiv preprint arXiv:1403.2877.

Thomsen, M., Moller, S. V. (2010). *Predictability of Long-Term Stock Returns* (master of art dissertation, Department of Business Studies, AARHUS University, UK)

Valizadeh, A. (2014). *Explaining a Model for Predicting Stock Returns*. (Doctoral dissertation), Alzahra University, Iran, Retrieved from http://simorgh.alzahra.ac.ir. (In Persian)

Wang, Y., & Choi, I. C. (2013). Market index and stock price direction prediction using machine learning techniques: an empirical study on the KOSPI and HSI. *arXiv preprint arXiv:1309.7119*.

Welch, I., & Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4), 1455-1508.

Westerlund, J., Narayan, P. K., & Zheng, X. (2015). Testing for stock return predictability in a large Chinese panel. *Emerging Markets Review*, 24, 81-100.

Zhang, L. (2002). *Essays on the Cross-Section of Returns*. (Doctoral dissertation, University of Pennsylvania, USA).

Zhou, Q., & Faff, R. (2017). The complementary role of cross-sectional and time-series information in forecasting stock returns. *Australian Journal of Management*, *42*(1), 113-139.