STATISTICAL INFERENCE ON SHARE MARKET: A REVIEW

Suni Ajaykumar*, Nagaraja Rao Chillale**

ABSTRACT

This paper is an attempt to study the importance and use of statistical techniques like testing of hypothesis, correlation and regression analysis etc in the study of stock market pattern. Random walk hypothesis, behavioral hypothesis, cross over effect, future equity returns, regression analysis of trading and stock returns are discussed here along with some data analysis of the study conducted using the historical data for S&P CNX NIFTY, the closing indices of the equity market from 1990 to 2008 the data of R2, volatility and returns of SENSEX scripts for one-year period of 30 companies of BSE. The statistical analysis for these data, using SPSS to emphasize the significance of statistical tools in the field of stock prices.

1. Introduction

Financial market, especially stock markets have grown considerably both in developed and developing countries over the last two decades. Increasing investor interest in emerging markets has motivated a great deal of research aimed at understanding the pattern and risk characteristics of stock prices in these markets. The investors could be expected to invest actively in productive enterprise only if

---

* Lecturer in Statistics, Kristu Jayanti College, Bangalore-560077
** Professor of Statistics, Vijaya College, Bangalore-560004

there was some mechanism by which they could sell a part of their stake in the business whenever they wished to generate cash. This need for making the investment 'liquid' was necessary to attract participation from individuals. Therefore the stakes of individuals in a business whether of ownership or of lending, had to be broken up into a large number of small units, so that each unit could be independently and easily bought and sold. This was achieved through shares and debentures representing smallest units of ownership and lending separately, by the public. Such breaking up of the stake into smaller denominations also helped in channeling small savings in the economy into entrepreneurial ventures. The institution where these are traded is known as the Stock Exchange. The movements of prices in the stock market are among a few phenomena that cut across the boundaries of academic disciplines and have cumulative research evidence spanning almost a century. Stock market has now become a crazy world. Different players enter the market with different time-frame intention. Some look for quick bucks, and would cash on the market momentum, totally ignoring the market and stock fundamental, while for some it is a casino, where they would make or lose money as if by a flip of coin. But, the investors will be benefited if they now the pattern of stock prices for a certain extent. Statistical techniques like time series, regression analysis, decision theory, probability theory estimation and hypothesis testing are very much useful in this field. The stock exchange is one of the most important institutions in the capital market, which includes term lending institutions, banks, investors, and just about anybody who are engaged in providing long term capital, to the industrial sector.

2. Hypothesis testing

Random walk hypothesis was first identified by French economist Louis Bachelier from the study of French commodity markets. Random walk hypothesis asserts that the random nature of the commodity of stock prices cannot reveal trends and therefore current prices are no guide to future prices. Random walk hypothesis is a financial theory states that market prices evolve according to a random walk and this cannot be predicted. Singal (2000) tested the random walk hypothesis and found that stock prices are less auto correlated subsequent to market opening. Although randomness of stock returns need not imply efficiency, it can generally claim that randomness is related to market efficiency. The random walk hypothesis is tested using the variance ratio test proposed by Lo and Mackinley (1988). Consider the recursive relation $X_t = \mu + X_{t-1} + \varepsilon_t$, Where $X_t = l_n P_t$ and P_t is the stock price at time t. The variable μ denotes draft, and ε_t is the random disturbance. The expectation of ε_t is assumed to be zero. Lo and Mackinlay (1988) develop two tests, a z-statistic under the assumption of homoskedasticity and a z* statistic that

allows for time varying volatilities. The hetroskedasticity-corrected z statistic is defined below.

$$Z^*(q) \;=\; \sqrt{nq}\, M_r(q) \,/\, \sqrt{\theta} \overset{o}{\sim} N\,(0,1);$$

where q is the number of overlapping base observations used for computing the variance, nq is the total number of base observations, and

$$\bar{M}_r(q) = \frac{\overline{\sigma_c^2(q)}}{\overline{\sigma_a^2(q)}} - 1 \quad ; \quad \overline{\sigma_a^2} = \frac{1}{nq-1}\sum_{k=1}^{nq}(X_k - X_{k-1} - \mu)^2$$

$$\overline{\sigma_c^2} = \frac{1}{m}\sum_{k=1}^{nq}(X_k - X_{k-1} - q\mu)^2$$

Alimov et al. (2004) examined stock prices on the BSE to test the random walk hypothesis

As mentioned earlier, random walk hypothesis can be tested using a variance ratio test that is

"Sensitive to correlated price changes but which is otherwise robust to many forms of hetropkedasticity and non-normality". The Lo & Mackinlay variance ratio test uses the fact that the variance of increments in a random walk is linear in the sampling period. This means the variance of $Y_t - Y_{t-2}$ is twice the variance of $Y_t - Y_{t-1}$. Consequently the plausibility of the random walk model by comparing the variance estimate of $Y_t - Y_{t-1}$ to, for example, the one-half or one fourth of $Y_t - Y_{t-2}$. Now consider the random walk process

$$Y_t \;=\; Y_{t-1} + \mu_t + \varepsilon_t; \qquad\qquad\qquad -----2$$

where Y_t is a series of stock returns of length $nq + 1$, μ_t is an arbitrary drift parameter and ε_t is a random disturbance term with zero mean and constant variance. They studied the random walk hypothesis using two statistical tests

i)   Dickey-Fuller Test for unit root

ii)   Lo and Mackinlay's variance ratio test

For the Dickey-Fuller test a regression equation is given by

$$Y \;=\; \rho\, Y_{t-1} + \varepsilon_t \quad ; \qquad\qquad\qquad ----\ (3)$$

where the ρ coefficient is the test for a unit root equal to 1. The term unit root refers to the root of the polynomial in the lag operator and implies the time series is nonstationary. If $\tilde{n}$ equals 1, and therefore the time series is a unit root, the stock can be thought of as following a random walk. For theoretical & empirical reasons, the Dickey-Fuller test is run as a regression

$$\Delta Y_t = (\rho - 1) Y_{t-1} + \varepsilon_t \quad \text{which is then rearranged as } \Delta Y_t = \delta Y_{t-1} + \varepsilon_t.$$

For Lo and Mackinlay's variance ratio test, the random disturbance term from equation (2) can either be assumed to have constant variance that is homoscedasticity, or time varying variance, that is hetrocedasticity. The variance ratio is

$$V(R (q)) \left[\sigma^2 (q) \ / \ \sigma(1) \right];$$

where

$$\sigma^2 (q) = \left[q \ (nq - q + 1) \ (1 - q / nq) \right]^{-1} \Sigma \left(Y_k - Y_{k-q} - q \, \tilde{y} \right)^2;$$

$$\sigma^2(1) = (nq - 1)^{-1} \sum_{k-=1}^{nq} (Y_k - X_{k-1} - \tilde{y})^2; \text{ and}$$

$$Y = (nq)^{-1}(Y(nq) - Y(0)).$$

In this case the test statistic is given by

$$Z_1(q) = \sqrt{nq} \left(VR(q) - 1 \right) (2(2q - 1)(q - 1)(q - 1) / 3q \text{ which has a zero mean and}$$
variance 1 for all step lengths of q. In the case of heteroscedastic error the test statistic is given by:

$$Z_2(q) = \sqrt{nq} \left(VR(q) - 1 \right) \left(V(q) \right)^{-1/2};$$

where

$$V(q) = \sum_{i=1} A(i)^2, \quad = \sum_{k=i-1} \frac{\left(Y_k - Y_{k-1} - Y \right)^2 \left(Y_{k-i} - Y_{k-i-1} - Y \right)^2}{\sum_{k=1}^{nq} \left(Y_k - Y_{k-1} - Y \right)^2}$$

and $A(j) = 2(q-j)/q.$

41

This statistic also has mean and variance 0 and 1 respectively. The authors gathered daily closing prices for the BSE 500 and BSE100 indexes and 14 stocks from July 4 2001 through October 31, 2003. The stocks on the BSE were numbered sequentially, and then random numbers were selected from a random number table to assemble the sample size.

The null hypothesis is that we do not reject that stocks follow random walk. Thus for the Dickey-Fuller test, $\delta = 0$. With a test statistic less than the critical value at an alpha level of 5 present, two-tail test and Lo and Mackinley's variance Ratio Test for the heteroscedastic error term of VR(q) = 1. The results of Dickey-Fuller test for unit root are shown in the following table. Observing each and every test statistic for the original time series of the BSE 500 index, BSE 100 index and all 14 stocks we find that they exhibit a unit root at the 5% alpha level. Therefore, the time series is non-stationary and the random walk hypothesis cannot be rejected.

Poshakwale (2002) also examines the random walk hypothesis in the emerging Indian stock market and makes some important contributions. He employs statistical tests capable of detecting linear as well as non-linear independence. The aim is constant that is testing for the random walk hypothesis; the absence of both linear and non-linear dependence should be confirmed since rejection of linear dependence doesn't imply independence but merely suggests a lack of linear auto correlation. They analyzes equally weighted portfolio of 100 stocks and a sample comprising of 38 of the most actively traded individual stocks and this is the first study that provides evidence on the return and risk characteristics of daily price stocks in the Indian stock market using a large set of data and a range of econometric tests.

The statistical evidence in this paper rejects the random walk hypothesis for the emerging Indian stock market. They used autoregressive model for removing linear dependence from daily return before investigating whether a potentially forecastable non-linear structure, non-linearity or other hidden patterns are present in linearly fitted returns. The test based on a null hypothesis of independent and identical distribution is widely used as a means of examining non-linear dependence and the adequacy of a variety of time series models. Truly random data will create a region of m space for any m deterministically generated data will only show geometric structure for sufficiently large m. Given a time series $\{Z_i : T = 1,.....T\}$ of D dimensional vectors, the correlation integral C(I) defined as

$$C(I) = \lim_{T \to x} \frac{2\sum_{i<j} I_i(z_i, z_j)}{T(T-1)} \; ;$$

where $I_l(x,y)$ is an indicator function that equals one if $\| x - y \| < 1$ and zero otherwise, where $\| \quad \|$ is the sup-norm. The correlation integral measures the fraction of pairs of points of $\{Z_t\}$ that are within a distance of l, from each other. The CD of $\{Z_t\}$ is thus defined as $\gamma = \lim \log C(l)$ if the limits exists.

$$l \to 0 \qquad \log l$$

If $\{Z_t\}$ were independently and identically distributed, then $C_m(l) = C_1(l)^m$.

Abraham (2005) also tested the random walk hypothesis for the Bombay Sensitive Stock Index (sensex). An important property of the random walk process is that the variance of the increments to the random walk process linearly increases with the sampling interval. If X_t is a pure random walk, the ratio of the variance of the qth difference scaled by q to the variance of the first difference must approach unity. The RWH for each of the markets is tested using the variance ratio test is computed for multiples of 2,4,8,16 periods, and the result indicates a rejection of random walk hypothesis at the 0.01 level.

Machael Stutzer develops behavioral hypothesis and provide a simple, distribution free approach to the estimate the optimal portfolio when the excess returns are not normally distributed. Denote a portfolio p's excess return in any time period t by R_{pt} and denote the time averaged excess return it earns over T periods by,

$$\overline{R}_{pT} \equiv \left(\Sigma \ R_{pt} \right) / \ T.$$

assuming the portfolio has a positive expected excess return, the law of large numbers implies that prob $\left(R_{pT} \le 0 \right) \to 0$ as $T \to \infty$. In i.i.d return processes and a wide variety of other return processes, this probability will eventually converge to zero asymptotically at a computable exponential rate,

$$I_p : \text{that is } pro\left(R_{pT} \le 0 \right) = c / \sqrt{T} e^{-I_p T}$$

For large T, where c is a constant that depends on the return distribution. Therefore the behavioral hypothesis is as follows.

A fund manager who is averse to receiving a non-positive time averaged excess return above some specified benchmark will direct analysts to select a portfolio m, that makes the probability of such a return occurring decay to zero at the maximum possible rate I_m. In computing excess returns, the manager will want to minimize.

Pro $\left(R_{pT} \leq 0\right)$ over an indefinite time span that is much longer than the contract interval.

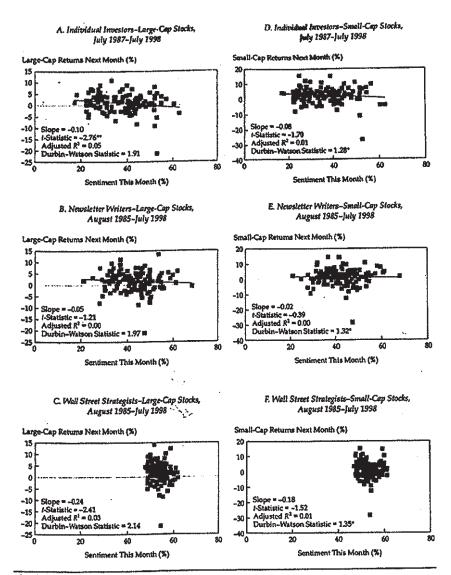# 3. Correlation and Regression Analysis

Fisher and Statman (2000) showed that the sentiment of Wall Street strategists is unrelated to the sentiment of individual investors or that of newsletter writers, although the sentiment of the last two groups is closely related. They found negative relationship between the sentiment of each of these groups and future stock returns. They studied three groups of investors- large, medium and small. The relationship between the sentiment of individual investors and the sentiment of newsletter writers is strong, whereas the relationship between the Wall Street strategists and the other two groups is not. Table 1 reports the changes in the sentiments of the three groups by month. As can be seen, individual investors grow bullish when newsletter writers grow bullish but not in lock step. The correlation of 0.47 is highly statistically significant but hardly perfect.

Table 1. Correlation between changes in sentiment

	Individual investors	Newsletter writers	Wall street strategists
Individual Investors	1	--	--
News letter writers	0.47	1	
Wall street Strategists	0.01	0.03	1

The changes in the sentiment of Wall Street strategists are virtually unrelated to changes in the sentiment of individual investors or newsletter writers.

Scatter plots for the relationship between the sentiments of the three groups and the returns of large-capitalization stocks and the relationships between the sentiments of the three groups and the returns of small-capitalization stocks are in Figure 1.

A. Individual Investors–Large-Cap Stocks, July 1987–July 1998

Large-Cap Returns Next Month (%)

Slope = –0.10
t-Statistic = –2.76**
Adjusted R^2 = 0.05
Durbin-Watson Statistic = 1.91

Sentiment This Month (%)

D. Individual Investors–Small-Cap Stocks, July 1987–July 1998

Small-Cap Returns Next Month (%)

Slope = –0.08
t-Statistic = –1.70
Adjusted R^2 = 0.01
Durbin-Watson Statistic = 1.28*

Sentiment This Month (%)

B. Newsletter Writers–Large-Cap Stocks, August 1985–July 1998

Large-Cap Returns Next Month (%)

Slope = –0.05
t-Statistic = –1.21
Adjusted R^2 = 0.00
Durbin-Watson Statistic = 1.97

Sentiment This Month (%)

E. Newsletter Writers–Small-Cap Stocks, August 1985–July 1998

Small-Cap Returns Next Month (%)

Slope = –0.02
t-Statistic = –0.39
Adjusted R^2 = 0.00
Durbin-Watson Statistic = 1.32*

Sentiment This Month (%)

C. Wall Street Strategists–Large-Cap Stocks, August 1985–July 1998

Large-Cap Returns Next Month (%)

Slope = –0.24
t-Statistic = –2.41
Adjusted R^2 = 0.03
Durbin-Watson Statistic = 2.14

Sentiment This Month (%)

F. Wall Street Strategists–Small-Cap Stocks, August 1985–July 1998

Small-Cap Returns Next Month (%)

Slope = –0.18
t-Statistic = –1.52
Adjusted R^2 = 0.01
Durbin-Watson Statistic = 1.35*

Sentiment This Month (%)

*Significant at the 5 percent level.
**Significant at the 1 percent level.

Figure 1: Relationships between Bullish Sentiment and Future Equity Returns

Panel A of Figure 1 shows that the level of sentiment of individual investors is reliable contrary indicator of future S&P 500 returns. There is a negative and

45

statistically significant relationship between the sentiment level of individual investors and S & P 500 returns in the following month. Similarly in Panel C we found a negative and statistically significant relationship between the sentiment level of Wall Street strategists and S & P 500 returns in the following month. Panel D, E and F of Figure 1 provide the scatter plots for the relationship of the level of sentiment with small-cap returns. In all these cases the level of sentiment did not forecast the returns of small-cap stocks as well as it forecasted the large-cap stocks.

Although there is a negative relationship between the level of sentiment and future small-cap stock returns for all these groups, this relationship was not statistically significant.

Moreover the low level of statistical significance in the small-cap stock regressions was biased upward by serial correlation in the residuals. Table 2 shows that a multiple regression of S & P 500 returns in one month on the level of sentiment of the three investor groups in the preceding month yielded on R2 of 0.08, a number that was statistically significant at the 1 percent level. Although an R2 of 0.08 might seem low because it indicates that sentiment explains an 8% of S & P 500 returns the 0.08 figure is properly interpreted as high.

Table 2. Multiple Regression of S&P 500 Returns on Level of Sentiment in Preceding Month

Dependent Variable		Independent Variable (July 1987–July 1998)		
Stock Returns Next Month	Intercept	Individual Investors	Newsletter Writers	Wall Street Strategists
Large stocks (S&P 500)	17.48	−0.09	−0.01	−0.24
t-Statistic		(−2.37)*	(−0.16)	(−2.43)*
Adjusted R^2				0.08
Durbin–Watson statistic				1.98
Small stocks (CRSP 9–10)	12.40	−0.08	0.00	−0.16
t-Statistic		(−1.45)	(−0.06)	(−1.25)
Adjusted R^2				0.01
Durbin–Watson statistic				1.28*

*Significant at the 5 percent level.

The authors found no signs of a meaningful relationship between change in sentiment in one month and stock returns in the following month. For large-cap stock returns, the relationship was negative for all sentiment groups but never statistically significant.

Peng and Seasholes analyzed the trading behavior of stock market investors and found that purchases and sales are highly correlated when we divide investors geographically. They examined account level data from the People's Republic of China (PRC). In the PRC brokerage rules require that an individual place all of his
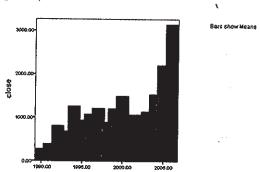
or her trades through the branch office where he or she opens the account. They had taken a sample, which consists of trades from seven branch offices. Four are located within the same province of the PRC and three are thousands of kilometers away in another municipality.

The layout of a typical PRC brokerage office allows for open conversation between investors. If investors engage in hers behavior due to animal spirits or group psychology, we would expect the trades of two isolated groups to be uncorrelated. But this study shows that isolated groups of investors engage in highly correlated trading behavior. The correlation of net trades is negative between two groups in different regions of the country.

4. Empirical study

The study conducted here uses the historical data for S&P CNX NIFTY, the closing indices of the equity market from 1990 to 2006 (http://www.nseindia.com) and conducted statistical analysis for these data using SPSS. The graph given below shows the interaction between the year and the closing value of the stock market indices.

Interactive Graph (Figure 2)

Bars show Means

An attempt is made to fit a suitable curve for the available data (the year and the closing value of the equity indices) and found that a linear relationship fits the given data approximately and the curve is also shown below which explains the results.
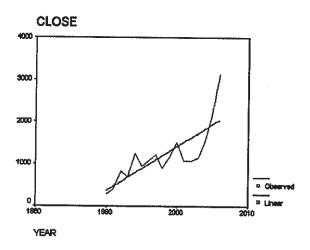
Curve Fit (Figure 3)

MODEL: MOD_1.

Independent: YEAR

Dependent Mth Rsq d.f.   F Sigf   b0   b1

CLOSE LIN .627   15 25.26 .000 -208385 104.899

CLOSE

Under the regression analysis, the following table (table 3) shows the strong relationship between the observed and predicted values of the dependent variable (the closing value).

Table 3: Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.792	.627	.603	421.54368

a Predictors: (Constant), YEAR

b Dependent Variable: CLOSE

In the ANOVA table (table 4) we can observe a small significance value of the F statistic, which indicates, that the independent variable do a good job in explaining the variation in the dependent variable.

48

Table 4: ANOVA

Model	Sum of Squares	df	Mean Square	F	Sig.
1	Regression 4489512.299	1	4489512.299	25.265	.000
	Residual 2665486.089	15	177699.073		
	Total 7154998.387	16			

a Predictors: (Constant), YEAR

b Dependent Variable: CLOSE

A run test has also been conducted to find out whether the closing values are random, and the large significant value in the run test table indicates that the values under test are random.

Table 5: Runs Test

	CLOSE
Test Value	1080.0000
Cases < Test Value	8
Cases >= Test Value	9
Total Cases	17
Number of Runs	8
Z	-.488
Asymp. Sig. (2-tailed)	.626
	a Median

The data of R2, volatility and Returns of SENSEX scripts for one-year period of 30 companies of Bombay Sock Exchange (http://www.bseindia.com) is used to conduct a statistical analysis of the data using SPSS and found that the correlation between beta and average volatility is considerably high and also the correlation between beta and R2 is also found to be significantly high.

Table 6: Correlations

		Beta	R2
Beta	Pearson Correlation	1	.801**
	Sig. (2-tailed)		.000
	N	30	30
R2	Pearson Correlation	.801**	1
	Sig. (2-tailed)	.000	
	N	30	30

**. Correlation is significant at the 0.01 level

Table 7: Correlations

		R2	avgvolatility
R2	Pearson Correlation	1	.600**
	Sig. (2-tailed)		.000
	N	30	30
avgvolatility	Pearson Correlation	.600**	1
	Sig. (2-tailed)	.000	
	N	30	30

**. Correlation is significant at the 0.01 level (2-tailed).

Table 8: ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	6399.490	1	6399.490	5.126	032(a)
	Residual	34959.057	28	1248.538		
	Total	41358.547	29			

a Predictors: (Constant), Beta
b Dependent Variable: returns

Table 9: Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	60.507	17.818		3.396	.002
	Beta	-38.112	16.834	-.393	-2.264	.032

a. Dependent Variable: returns

Table 9 and table 10 show that there is no significance difference of the average beta value from 1 at 5% level of significance.

Table 10: One Sample Test

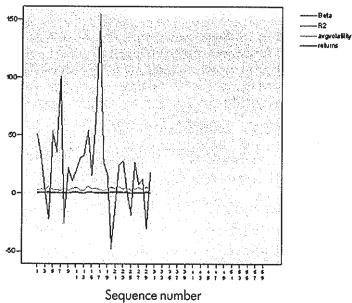|	Test Value = 1					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Beta	-.187	29	.853	-.01333	-.1589	.1322

Sequence number

Figure 4.

51

The figure above (Figure 4) shows the variations in the values of beta, R^2, average volatility and returns.

5. Discussion

Stock market prediction is the act of trying to determine the future value of a company stock traded on a financial exchange. The successful prediction of a stock's future price could yield significant profit. Some believe that stock price movements are governed by the random walk hypothesis and thus are unpredictable. Others disagree and those with this viewpoint possess a myriad of methods and technologies which allow them to gain future price information. In this article an attempt is made to discuss the different statistical methods used in the field of stock market analysis. In this paper an attempt has been made to study and understand various statistical methods used in this field such as testing of hypothesis, regression analysis etc. In the empirical study conducted here, some simple and familiar statistical methods are used to study the data using the statistical software package SPSS and concluded the study by emphasizing the important role of Statistics in the field of stock market analysis.

Even though many useful statistical methods in the field of stock market analysis are discussed, empirical study has not conducted for all those methods. It is likely that the quest for the methods of analysis of the Stock Market will continue to stimulate the intellect of academic researchers. It is apparent that over the years, the field has made much progress and that without such sustained research efforts the issues in this field would remain unresolved. A study on the behavior of volume and stock return around record date can be conducted using statistical tools. This article presents many statistical models to simulate the dynamics of a stock market.

References

1. Michael Stutzer,"A portfolio Performance Index",Financial analysysts journal.May/June(2000).

2. Sunil Poshakwale: "The random walk hypothesis in the Emerging Indian Stock Market."*Journal of Business Finance and Accounting.* November/December (2002).

3. Azizjon A.Alimov, Debashish Chakkraborty, Raymond. A.K Cox, Adishwar K Jain: "The random walk hypothesis on Bombay Stock Exchange", *Finance India,* Volume 18 No 3, September (2004).

4. Abraham Abraham: "Non-Synchronous trading and the investigation of market Efficiency-The Bombay Stock Index", *The ICFAI Journal of Applied Economics,* November (2005).

5. Kenneth I. Fisher and Meir Statman: "Investor sentiment and Stock Returns", *Financial Analysts Journal*. March /April (2000).

6. Lei Peng and Mark S Seasholes: "Correlate Trading and Location", *The Journal of Finance*, October (2004)

7. Vijay Singal: "Stock Market Openings: Experience of Emerging Economies", *Journal of Business* (2000).

8. Tano Santos Pietro Veronesi: "Labor Income and Predictable Stock Returns", *A Review of Financial Studies* (2006).