

IJEAST

INTERNATIONAL JOURNAL
OF ENGINEERING APPLIED SCIENCE
AND TECHNOLOGY

VOLUME : 4 ISSUE : 11 Print / Issue Publication Date: 10-May-2020

ISSN : 2455-2143

DOI : 10.33564/IJEAST.2020.v04i11.023

Indexed In

Google SCHOLAR OPEN ACCESS doi Crossref

WWW.IJEAST.COM

editor@ijeast.com

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 4, Issue 11, ISSN No. 2455-2143, Pages 123-127
Published Online March 2020 in IJEAST (http://www.ijeast.com)



SML ON HIGH FREQUENCY TRADING DATA

Harshit Joshi
Cluster Innovation Centre
University of Delhi, New Delhi, India

***Abstract*— Market traders often try to predict the financial market trends and try to extract information from the trading of the various securities. To extrapolate these trends, practitioners often use a fundamental concept of capital asset pricing model (CAPM). The main purpose of this paper is to explore the High-Frequency Trading (HFT) data to verify the Efficient Market Hypothesis (EMH) as assumed by the Security Market Line (SML) using statistical methods and to find abnormal returns above the market, if exist, using the proposed model. The paper then validates the pragmatism of the obtained abnormal return using a rolling window of short interval of time trading strategy.**

Keywords— ***High Frequency Trading; Security Market Line; Abnormal Returns; CAPM***

I. INTRODUCTION

Market traders often try to predict the financial market trends and extract information from the trading of the various securities. The time series regression of security returns is often used to study the abnormal returns of an asset but it is used over daily returns and using a rolling window of short interval of time. However, the monthly estimates can be noisy and might not capture all the information from the assets pricing. One of the commonly used models in financial economics is the Capital Asset Pricing Model (CAPM) [Sharpe (1964), Litner, (1965), Jan Mossin, (1966)]. CAPM is a linear model that describes the relationship between non-diversifiable risk, also known as systematic risk or market risk, and the expected return for assets. CAPM was a landmark in the history of financial management, researchers and practitioners have applied CAPM to understand the market and to make decisions based on the returns of the asset but most of the study has been done on time periods as long as a couple of years or at least a few weeks. This risk is the ratio of covariance of market return and asset return to the volatility of the market return. The linear models such as CAPM are not able to truly explain the sudden and unscheduled changes in the market.

To tackle the problem of noisy data from the daily return and to explain the sudden changes in the asset pricing, this paper turns to the High-Frequency Trading (HFT), which a computer-based trading method and algorithms depend on the execution time and the pricing of the assets without any human interaction. For individual assets, we use Security Market Line (SML) and its relation to expected return and non-diversifiable risk to verify the Efficient Market Hypothesis (EMH) an idea which goes back to Bachelier (1990), Mandelbrot (1963), Samuelson (1965) and to find, if exists, abnormal returns above the market return using statistical methods.

There are other models based on CAPM which are commonly called factor models, also known as Fama-French factors [Fama et al. (1992, 1993, 2015)] but the models try to explain different kinds of factors on which the expected return of the asset depends on. The main purpose of this paper is to find if there exists abnormal return, if exists, using the one-factor model, where the factor is market premium. The abnormal returns are then hypothesised using statistical methods and over different short intervals of time. In the next section of this paper in Section II describes the model used and the approach taken by this paper followed by describing the data used for the modelling and visualization, to find abnormal returns above the market. The returns are then studied over different time intervals and verified using a rolling window technique and the results are printed in Section III. Finally, in Section IV, we discuss the results presented in this paper and the future approaches that one can take.

II. DATA AND METHODOLOGY

The main purpose of this research is to verify the EMH assumed by the SML, used for individual assets in the financial market, with HFT and to find if there exist abnormal returns above the market which can be used to take a long or short position. The chosen model is fit to the HFT data and the paper then tries to explain the behaviour of abnormal returns and the systematic risk using frequency data, ranging from 1 minute to 10 minutes.

A. Data

The data includes trades and quotes data of 15 stocks from The National Stock Exchange Fifty (NIFTY 50) index. NIFTY 50 represents the weighted average of 50 Indian stocks in 13 sectors. The data is a 4-day intra-day dataset of 15 securities and NIFTY 50 index. All are highly traded stocks and are used for the same purpose, the 15 securities represent all the 13 NIFTY 50 sectors. For time series regression of equation 3 on the data, the paper uses 11 securities along with NIFTY 50 index. Out of the 15 securities in the data provided, 4 of them had discrepancies and hence weren't used.

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 4, Issue 11, ISSN No. 2455-2143, Pages 123-127
Published Online March 2020 in IJEAST (http://www.ijeast.com)

Table 1 lists all the securities used for the simulation along with their then market cap. The stocks eliminated from the list are Wockhardt Ltd., Mindtree Ltd., Tata Global Beverages Ltd. and CESC Ltd.

Table 1: List of Securities used for simulation

FIRMS	MARKET CAP (INR Cr.)
TCS	490390.25
HINDUNILVR	171789.53
TATAMOTORS	109338.40
POWERGRID	69998.66
HEROMOTOCO	52159.37
GAIL	41320.65
YESBANK	32279.17
OIL	26627.31
TATASTEEL	24124.99
BANKINDIA	11229.32
APOLLOTYRE	9864.90

The tick-frequency data was converted to time-frequency data and then windowed to indicate prices at 1 minute, 2 minutes, 3 minutes, 5 minutes, 6 minutes and 10 minutes time intervals. In the model the only concerne is the price of the security hence, we keep the last traded price in our data.

B. Model

According to the EMH, in an efficient market at any point in time, the actual price of a security will be a good estimate of its intrinsic value [Fama (1970)]. EMH means that asset prices reflect all the available information and it is impossible to "beat the market" consistently on a risk-adjusted basis since market prices should only react to new information. SML assumes that there exists an efficient market and only one factor is considered. The SML is given by equation

$$\mathbb{E}(\boldsymbol{R_i}) = \boldsymbol{R_f} + \boldsymbol{\beta_i}\big(\mathbb{E}(\boldsymbol{R_m}) - \boldsymbol{R_f}\big) \qquad (\mathbf{1})$$

Here, $E(R_i)$ is the expected return on the capital asset, R_f is the risk-free rate of interest such as interest arising from government bonds, β_i (the beta) is the sensitivity of the expected excess asset returns to the expected excess market returns, or also, $E(R_m)$ is the expected return of the market, $E(R_m) - R_f$ is sometimes known as the market premium, the difference between the expected market rate of return and the risk-free rate of return.

The $E(R_i) - R_f$ is known as the risk premium and through the readjustment of variables, we can show that the equation (1) can be written as

$$\mathbb{E}(R_i) - R_f = \beta_i\big(\mathbb{E}(R_m) - R_f\big) \qquad (2)$$

To check for if the SML correctly fits the data, we can use a linear model with an additional variable which takes in to account the unexplainable part of the SML. So, for the modelling of the HFT data, we use the equation (3)

$$\mathbb{E}(R_i) - R_f = \alpha_i + \beta_i\big(\mathbb{E}(R_m) - R_f\big) \qquad (3)$$

Upon regressing the HFT data using equation 3, if the EHM holds, then the value of α_i should be zero else we will find some significant value of α_i based on the p-value from statistics. Later in the paper, we will discuss the significance of this alpha and how can we take advantage of knowing the value of this alpha (α_i).

For simulation, we use log-return of NIFTY50 and the securities. log-return is the consecutive differences in prices in log-scale, i.e.,

$$r_t = \log(P_t) - \log(P_{t-1}) \qquad (4)$$

$$= \log\left(\frac{P_t}{P_{t-1}}\right) \qquad (5)$$

Where P_t is the price of the stock (or the value of the index) at the timestamp t. Using equation (5) the model (3) becomes:

$$r_i - R_f = \alpha_i + \beta_i\big(r_m - R_f\big) \qquad (6)$$

$$R_f = \frac{R_{f(annualised)}}{250 * l_i} \qquad (7)$$

Where l_i is the number of timestamps per day, r_m is the log return of market return and r_i is the log return of the assets.

C. Simulation

We used the equation 6 to perform linear regression to find the coefficient values of α_i and β_i taking using 6 different time-frequency data, 1 minute, 2 minutes, 3 minutes, 5 minutes, 6 minutes and 10 minutes. For the simulation purposes, annualised risk-free return ($R_{f(annualised)}$) was taken as 6.0.

III. Empirical findings

Upon applying the linear regression on the HFT dataset we were able to find a few interesting results. In this section, we will go through various such findings and will try to interpret them.

Alpha α and Beta β are the two coefficients we tried to estimate while using equation 6. For each of these coefficients, we also find t-statistics and p-value to reject or accept the null hypothesis that the value of these coefficients is zero. Table 2 shows that when the time interval is less than 180 seconds, according to the p-value, alpha cannot be ignored despite being very small. This pattern is not restricted to TCS only, in figure 2 we can see that when the time interval is less than 180 seconds, the p-value of α reduces with fewer outliers whereas according to the p-value, we can say that the non-diversifiable risk $\beta_i(r_m - R_f)$ for an individual asset is 0 when the time-frequency is less than 60 seconds, this gives another research problem to investigate the beta (β_i) for HFT. It has been previously studied with 5 minutes time-frequency data that the β_i is significant in Tsay et al. (2004) but the paper doesn't explore this with smaller time intervals.

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 4, Issue 11, ISSN No. 2455-2143, Pages 123-127
Published Online March 2020 in IJEAST (http://www.ijeast.com)

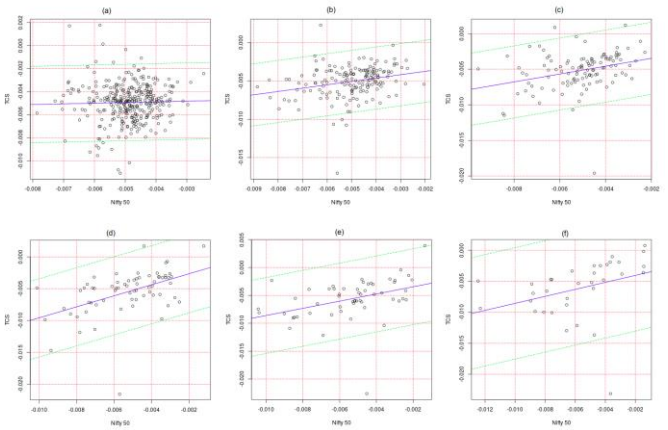


Fig. 1. Linear Regression across different time-frequency. (a - f) correspond to 1 minute, 2 minutes, 3 minutes, 5 minutes, 6 minutes and 10 minutes data frequency

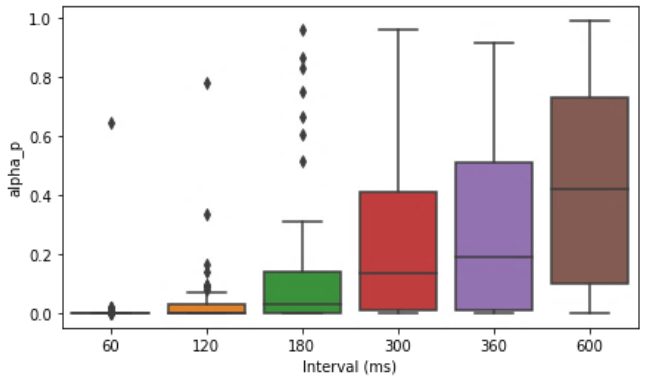


Fig. 2. p-value of alpha vs data frequency

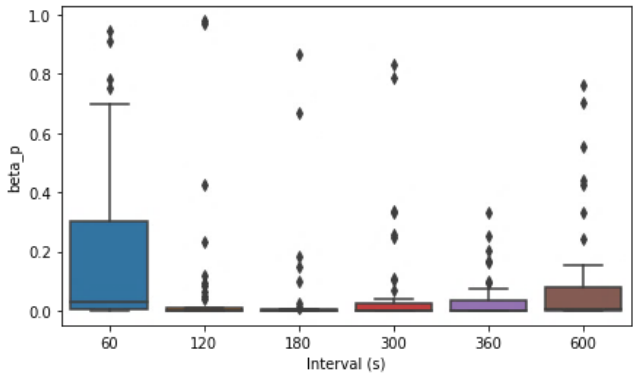


Fig. 3. p-value of beta vs data frequency

A few relations which looked interesting to us were the relationship of alpha with adjusted R squared. The figure 4 and 5, show that the adjusted R squared for the linear model (equation 6), is smaller than 0.2 for smaller time interval and when the coefficient (α), abnormal return, is significant enough.

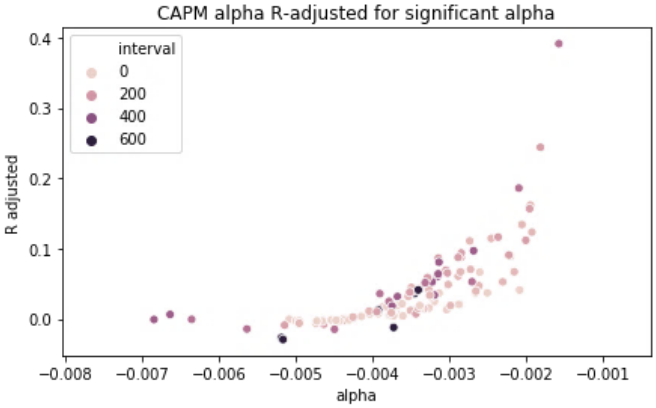


Fig. 4. Adjusted R sq. vs alpha values for p-value < 0.05

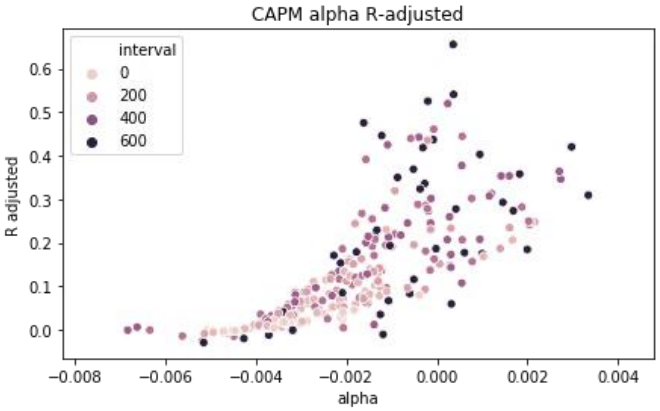


Fig. 5. Adjusted R sq. vs alpha values

IV. RESULTS AND DISCUSSIONS

A. Interpretation

Since we have to discard our null hypothesis that the value of α, abnormal return, is 0 we can say that there are other factors apart from the market premium which affects the risk premium and there is extra information which isn't explained by the SML. Using this abnormal return above the market we can say that, α is a coefficient which indicates whether the asset is undervalued or overvalued according to its current price. According to the Arbitrage Pricing Theory (APT) [Ross (1976)] if a security is mispriced then the market eventually corrects itself and the security moves back to its fair price. If $\alpha_i > 0$ this indicates that the security is undervalued and the stock price will rise to its fair pricing whereas when $\alpha_i < 0$, it indicates that the security is undervalued and the stock prices will fall to its fair price. When $\alpha_i = 0$, it indicates that the value of the stock is its intrinsic value, equation 3 becomes equation 2 and we can say that Efficient Market Hypothesis holds true and all the securities are priced fairly.

B. Trading Strategy

To verify the pragmatism of alpha or abnormal returns that we get from the model, we use a trading strategy. The trading

International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 4, Issue 11, ISSN No. 2455-2143, Pages 123-127
Published Online March 2020 in IJEAST (http://www.ijeast.com)

strategy proposed in this paper depends on the value of alpha α_i, and then taking a long or short position. The algorithm used is as follows.

1. Take log return of asset return and market return.
2. Use the linear model proposed for a 60-minute window time interval using 1-minute time-frequency.
3. If alpha is greater than 0 then take a long position else take a short position and calculate gain.

Algorithm 1: Trading Strategy
Input: stock
Output: Total Gain

1. n = length(stock)
2. interval = 60
3. $R_{f(annualised)}$ = 6.0
4. $R_f = \frac{R_{f(annualised)}}{250*n}$
5. **for** i = 1 **to** (number of windows):
6. $\log R_i$ = diff(log(stock[window]) - R_f
7. $\log R_m$ = diff(log(nifty[window]) - R_f
8. capm_stock = lm($\log R_i$, $\log R_m$)
9. CALCULATEGAIN(capm_stock)
10. **end for**

Gains can be calculated in two ways; first one (Algorithm 2) is a simple gain calculation which penalised if the short position is taken but the stock price rises whereas the second method (Algorithm 3) focuses completely on the money made and doesn't penalise for taking short position.

Algorithm 2: Calculate penalized gain
Input: capm_stock
Output: Calculated gain

1. **if** α_i >= 0
2. window_gain = stock[window + 1] – stock[window]
3. **else if** α_i < 0
4. window_gain = stock[window] – stock[window + 1]
5. **end if**

Algorithm 3: Calculate non-penalized gain
Input: capm_stock
Output: Calculated gain

1. **if** α_i >= 0
2. window_gain = stock[window + 1] – stock[window]
3. **else if** α_i < 0
4. window_gain = MAX(stock[window] – stock[window + 1], 0)
5. **end if**

Table 2 and 3 show the daily gains per stock according to Algorithm 2 and 3 respectively.

Table 2: (a) Daily gains of stocks according to algorithm 2 (b) points per stocks for 4 days in INR.

(a)

Stock	Day 1	Day 2	Day 3	Day 4
TCS	Gain		Gain	
HINDUNILVR		Gain		Gain
TATAMOTORS	Gain	Gain		
POWERGRID	Gain	Gain	Gain	
HEROMOTOCO		Gain	Gain	Gain
GAIL				
YESBANK	Gain		Gain	
OIL				
TATASTEEL	Gain	Gain	Gain	
BANKINDIA				Gain
APOLLOTYRE	Gain		Gain	

(b)

1.10	-18.00	-18.70	2.95	43.75	-19.05
-0.15	-109.15	27.75	-35.40	-3.95	

Table 3: (a) Daily gains of stocks according to algorithm 3 (b) point per stocks for 4 days in INR.

(a)

Stock	Day 1	Day 2	Day 3	Day 4
TCS	Gain	Gain	Gain	Gain
HINDUNILVR	Gain	Gain	Gain	Gain
TATAMOTORS	Gain	Gain	Gain	Gain
POWERGRID	Gain	Gain	Gain	Gain
HEROMOTOCO	Gain	Gain	Gain	Gain
GAIL	Gain	Gain	Gain	Gain
YESBANK	Gain	Gain	Gain	Gain
OIL	Gain	Gain	Gain	Gain
TATASTEEL	Gain	Gain	Gain	Gain
BANKINDIA	Gain	Gain	Gain	Gain
APOLLOTYRE	Gain	Gain	Gain	Gain

(b)

200.50	58.65	226.15	26.70	465.85	90.85
72.25	183.75	169.10	87.35	11.50	

Using algorithm 2 to calculate gain and validate the model, we obtain positive returns in 20 out of 44 slots, whereas using algorithm 3 we saw a positive return on investment in 44 out of 44 slots. The results according to table 2 and 3 shows that the positive stock movement is correctly captured by the model as proposed in Section II, whereas the model doesn't explain the negative movement of the stock.

V. CONCLUSION

After studying the 4-day High-Frequency Trading data for 11 securities we can say that, when the stocks are traded with small time-frequency, they react less to how the market behaves, figure 3 shows that for time interval less than 1 minute we accept

the null hypothesis that the coefficient $\beta = 0$ based on the statistical method as mentioned in section II. From figure 2, we observe that as the time-frequency for observed prices increases the p-value increases and we accept the null hypothesis, $\alpha = 0$, for time-frequency greater than 3 minutes. We can also interpret that we can take advantage of the market when the frequency of data is high, less than 3 minutes, as the stocks give abnormal returns and don't react to market or outside news. The other hypothesis from this paper can be taken that the time taken by external information to influence a particular stock is less than 1 minute.

The trading strategy given in section III to validate the significance of alpha shows that the trader can take the advantage of the market in High Frequency Trading environment and that the model proposed captures the positive movement of the stocks but fails to capture their downward movement.

Future work that can follow this paper is working with larger HFT dataset to further authenticate the behaviour of individual stocks in HFT. Using Factor models as proposed by (Fama and French, 1992, Fama and French, 1993 and Fama and French, 2015) can be used to further study these securities.

VI. ACKNOWLEDGEMENT

The data sponsored by Finance Research Group, Indira Gandhi Institute of Development Research during Mathematical Finance Summer School 2019 at Chennai Mathematical Institute.

We thank our colleagues from Cluster Innovation Centre who provided insight and expertise that greatly assisted the research, although they may not agree with all of the interpretations/conclusions of this paper.

VII. REFERENCE

[1] Engle R.F. (2000), The Econometrics of Ultra-high-frequency Data. Econometrica, 68: (Pg 1-22).

[2] Fama F. Eugene, French K. R. (2015). A five-factor asset pricing model. Journal of Financial Economics, 116(1), (Pg1-22).

[3] Fama F. Eugene, French K. R. (1993). Common risk factors in the returns on stocks and bonds, Journal of Financial Economics, 33(1), (Pg3-56).

[4] Fama F. Eugene and French K.R. (1992), The Cross-Section of Expected Stock Returns. The Journal of Finance, 47: (Pg427-465).

[5] Fama F. Eugene (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. The Journal of Finance, 25(2), (Pg383-417).

[6] Lintner, J. (1965). The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. The Review of Economics and Statistics, 47(1), (Pg13-37).

[7] Mossin, J. (1966). Equilibrium in a Capital Asset Market. Econometrica, 34(4), (Pg768-783).

[8] Ross, S. A. (1976). The arbitrage theory of capital asset pricing. Journal of Economic Theory, 13(3), (Pg341–360).

[9] Sharpe, W.F. (1964), Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk. The Journal of Finance, 19: (Pg425-442).

[10] Tsay, Ruey. (2003). Non-synchronous Trading and High-Frequency Beta.

[11] Bachelier, L. (1900). "Théorie de la spéculation". Annales Scientifiques de l'École Normale Supérieure. 17: (Pg21-86).

[12] Mandelbrot, Benoit (1963). "The Variation of Certain Speculative Prices". The Journal of Business. 36 (4): (Pg394).

[13] Samuelson, Paul A. (2015), "Proof that Properly Anticipated Prices Fluctuate Randomly", The World Scientific Handbook of Futures Markets, World Scientific Handbook in Financial Economics Series, **5**, WORLD SCIENTIFIC, (Pg25-38).

[14] "Broadmarket indices - NIFTY 50 Index". nseindia.com. NSE - National stock exchange (official website)

IJEAST

INTERNATIONAL JOURNAL OF ENGINEERING APPLIED SCIENCE AND TECHNOLOGY

AUTHORS

Harshit Joshi

ABOUT IJEAST

International journal of Engineering Applied Science and Technology (IJEAST) is a peer-reviewed, open access journal that publishes high - quality research paper in the field of Engineering, Applied Science and Technology.

IJEAST aims to provide a platform for researchers, academicians, and professionals to share their innovative ideas, research findings, and practical experiences with the global scientific community.

JOURNAL METRICS

H-INDEX	i10-INDEX	IJEAST CITATIONS	PEER-REVIEWED & OPEN ACCESS
33	154	9,647 (IJEAST CITATIONS)	PUBLISHED MONTHLY

PEER REVIEWED

All submission are rigorously peer reviewed to ensure quality.

AUTHORS ARE OUR PRIORITY

We value our authors and recognize their contribution to the advancement of research and knowledge.

OPEN ACCESS

Free and unrestricted access to research for all.

GLOBAL REACH

Connecting researchers and professionals worldwide.

TIMELY PUBLICATION

We ensure a swift and efficient publication process.

For more information, visit our website
www.ijeast.com

editor@ijeast.com
www.ijeast.com
India

2455-2143