

# Applying the Support Vector Machine for Testing Pricing Inefficiency on the Stock Exchange of Mauritius

Aleesha Mohamudally-Boolaky<sup>1</sup>, Teemulsingh Luchowa<sup>1</sup> & Kesseven Padachi<sup>1</sup>

<sup>1</sup> University of Technology, Mauritius

Correspondence: Aleesha Mohamudally-Boolaky, University of Technology, Mauritius.

Received: July 3, 2019

Accepted: August 22, 2019

Available online: August 29, 2019

doi:10.11114/aef.v6i5.4495

URL: <https://doi.org/10.11114/aef.v6i5.4495>

## Abstract

A popular Machine Learning Technique called the Support Vector Machine (SVM) is adopted on the Stock Exchange of Mauritius (SEM) to determine if stock market returns are predictable based on information from past prices, allowing arbitrage opportunities for abnormal profit generation. The serial correlation test, used as benchmark, and the SVM technique show evidence that previous information on share prices as well as the indicators constructed are useful in predicting share price movements. The implications of the study are that investors have the prospect of adopting speculative strategies and profits from trading based on information and advanced techniques and models are possible.

**Keywords:** support vector machine, arbitrage pricing theory

**JEL classification:** C14, D53, G14

## 1. Introduction

Predictability of returns on any asset has been a very debatable topic in finance. A large strand of literature is dedicated to the subject, with varying results using different models. Recently, with the explosion in the development of new technological tools, Machine Learning Techniques have become even more accessible coupled with reduced cost of existing tools related to trend analysis.

The Efficient Market Hypothesis (EMH) advocated by Fama (1970) posits that past information is already incorporated into the current prices of assets and hence cannot be used as a basis for predicting stock trends. An instinctive and simple test for EMH is looking at whether the price of a stock is a function of its prices at previous periods. However, techniques in revealing patterns in past price information have evolved over time. While the prices may not be good indicators of future prices, market traders now use more complex methodologies to study movements and deviations so as to adopt different strategies on stock markets for short selling and arbitrage generally (Kim, 2003; Ni, Ni & Gao, 2011; Rosillo, Giner, Puente, & Ponte, 2013 and Rosillo, Giner & De la Fuente, 2014). The use of some of these techniques, has been proposed in this study as a better way in studying the predictability power of trends from past stock returns rather than just the returns themselves.

A different approach, being classification models, can instead be used to classify the expected return (for example, whether they will be positive or negative) based on the results of the methodologies. With advancement in technology and scientific methods, various machine learning methods are being put forward as efficient classification techniques in identifying trends from previous data. These machine learning techniques may not require checking the linearity assumptions inherent to ordinary least squares while also having more flexibility in adapting to new information as such models can identify insights from data and/or can be trained. Some of these techniques include Artificial Neural Networks(ANN), Fuzzy logic, Support Vector Machines (SVM), particle swarm optimisation and others (Sheta, Ahmed, & Faris, 2015). For this study, we use SVM classification method as an alternative to ordinary least squares regression analysis in an attempt to get the best insights from the methodologies put forward in this study. *“Support vector machines is a family of algorithms that have been implemented in classification, recognition, regression and time series.”*, Kara, Boyacioglu & Baykan (2011). The model was introduced by Vapnik (1999) and is a widely-used machine learning technique in a variety of applications such as prediction of Stock prices, image classification, text categorisation, among others. SVM’s advantage over regression analysis, for the purpose of this study, is that it can efficiently perform non-linear classifications. Accordingly, assumptions related to linear separation of data is not necessary. While a nonlinear transformation is used as part of the process, the transformation is based on robust

theoretical basis. In the real world, the data under study is not expected to follow a linear trend.

This present study focuses on presence of the weak form of EMH on the Stock Exchange of Mauritius (SEM) which was setup in 1989 and has evolved into a diversified market with 56 listed companies and 91 listed securities, comprising of ordinary shares and preference shares, debt securities, Exchange Traded Funds and global securities. The market capitalisation by industry as at 2016, as represented on the Official Market, is as per Figure 1. Previous studies (Bundoo, 2000 and Fowdar Subadar, Lamport, Sannasee, & Fawzee, 2007) on the SEM have found anomalies and evidence that the stock market is not efficient, suggesting predictability of Stock returns being possible. Given the empirical findings of the studies of the SEM not being in line with the EMH, the objective of this paper is to contribute in testing whether the inefficient state of the market indicated by previous research persists or there is rather evidence of an efficient market, in the weak form of EMH. The study also consequently determines if stock market returns are potentially predictable based on information from past prices, allowing arbitrage opportunities for abnormal profit generation using the SVM. To the best of our knowledge the only substantive study that has been conducted on companies listed on SEM using the Machine Learning Technique is the one by Cheeneebash, Galanis, Gopaul & Bhuruth (2008). They used the Artificial Neural Networks (ANN) as a forecasting technique for four leading companies operating in Mauritius. However, recent studies by Shen, Jiang & Zhang (2012), Wang (2014) and Sheta et al. (2015) suggest that the SVM appears to be a more accurate technique than the ANN for forecasting financial time series. Furthermore, Cheeneebash et al. (2008) focuses their study only on 4 companies, which would not be representative of the population on SEM. In an attempt to see whether trend analysis based on the daily returns are superior indicators of future returns and provide insight of the validity of EMH on the SEM, the SVM classification model is used. As such, the SVM classification method is adopted as an alternative to both the serial correlation test and ordinary least squares regression method for ascertaining EMH and getting improved insights from techniques generally used by market traders instead of studying only temporal variations in prices of a stock. The sample in the present study also covers a larger proportion of the companies listed on SEM. To our knowledge, the trend indicators used in this study, have not been tested to the same extent as in other studies involving price information from companies listed on the SEM.

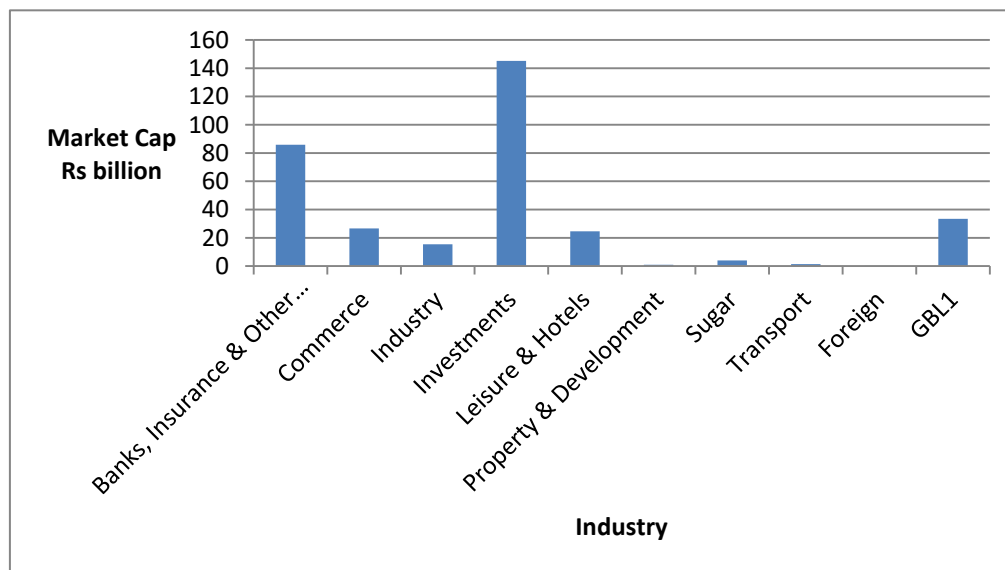


Figure 1. Market Capitalisation on the SEM Official Market (2016)

Source: SEM Factbook 2017

The rest of the paper is organised as follows: Section 2 gives an overview of the theoretical and empirical findings. The methodologies adopted, are described in Section 3 while the findings are discussed in Section 4. Conclusion and implications of study appears in Section 5.

## 2. Literature Review and Empirical Findings

### 2.1 The Efficient Market Hypothesis

Fama (1970) introduced three types of tests for market efficiency, as follows: The weak form of market efficiency is mainly based on the Random-Walk theory as analysed by Fama (1965). Since the Random-Walk theory assumes that the price of a Stock already incorporates all information in past prices, sequences of price fluctuations and other

information available in the market (even anticipated information), investors cannot use past information on Stock prices to formulate trading strategies with the purpose of earning profits. The semi-strong form is related to the speed to which Stock prices adjust to announcement of new information which is publicly available, for example annual reports, industry surveys, new security issues, dividend pay-out and other corporate actions announcements. Consistent with Fama's (1965) research and evidenced by other empirical results presented in his paper, these announcements would normally be incorporated in the price of a Stock instantaneously or not have any significant effect on share price fluctuations, with any bubbles being corrected by market forces. The strong form test concerns a situation where certain investors or groups of persons would have monopolistic access to information that allows them to earn abnormal profits on the stock market.

### *2.2 Tests of the Weak Form of EMH*

Kim & Shamsuddin (2008) performed measures of return predictability using automatic variance ratio test, automatic portmanteau test and generalised spectral test on the Dow Jones Industrial Average index over the period from January 1900 to June 2009. They found that return predictability fluctuates over time, governed largely by changing market conditions. Kim & Shamsuddin (2008) performed multiple variance ratio tests of Random-Walk based on wild bootstrap and signs, on nine Asian stock markets: Hong Kong, Japan, Korea, Indonesia, Malaysia, Philippines, Singapore, Taiwan and Thailand. The authors concluded that the weak form of efficiency is present on the developed or advanced emerging markets, which are Hong Kong, Japan, Korea, Singapore, Taiwan. The secondary emerging markets, Indonesia, Malaysia and Philippines, demonstrate market inefficiency. Anghel (2015) did a very comprehensive test of stock market efficiency in the weak-form on stock markets of 75 countries around the world, with 1,336 companies selected in the sample, during the period from January 2001 to December 2012 by using basic trading simulation based on the moving average convergence divergence ("MACD") indicator. The author found that countries demonstrating the highest market efficiency were Colombia, followed by Australia and Austria while the least efficient ones being Serbia, Vietnam and Ukraine. Overall, 34 inefficient markets were detected. Also, by aggregating the results year-wise, the author found that crises had a negative effect on market efficiency, most notably in 2008 and 2011.

In an assessment of the SEM, Bundoo (2000) performed a test of the weak-form efficiency using the serial correlation test to find if share prices on the exchange exhibit positive or negative correlations for the period during the years 1991 to 1999, based on the argument that only a zero correlation would be consistent with the weak form of the EMH. His results show a strong positive correlation in returns of the Stocks indicating market inefficiency in the weak-form. Fowdar et al. (2007) also performed an autocorrelation test which revealed that daily Stock returns were serially correlated. The Stock returns seemed predictable and reacted very slowly to new information.

Pan (2014) used SVM as both a classification model and a regression model to study historical prices of Google Inc. for the periods from 2004 to 2013 and was able to obtain prediction accuracy between 62% and 93.5% using the classification model and between 61% and 70% using the regression model. This suggests that the classification model was a better method than the regression variation of the SVM. Wang (2014) used the SVM to try to forecast movements in direction of The Korean Composite Stock Prices Index 200 ("KOSPI") and the Hang Seng Index ("HSI") and their constituents for the period from 2002 to 2011. For forecasting movements in stock prices of the indices, daily lagged prices for the indices and overall constituents are used as internal factors while for forecasting movements in stock price of the constituents, the relevant index and daily lagged prices of the specific constituent are used. The S&P 500 index and the exchange rate of US dollars to the Korean Won or Hong Kong Dollar are used as appropriate. Artificial Neural Networks ("ANN") and random walk test are also done to test the performance of the results from SVM. The results show a hit ratio varying between 50% and 65% for forecasting the movements in the indices with the SVM being minorly better than the ANN method. However, high hit ratios are achieved for prediction of movement in direction of the constituents of the indices, hitting range above 55% and below 70% for the KOSPI and between 55% and 63% for HSI. Patil, Patidar & Jain (2016) performed a SVM classification analysis over the years from 2004 to 2015 to try to predict movements of prices on a random selection of 13 companies from Yahoo Finance. The SVM model correctly classified 10 stocks out of the sample of 13 with respect to predicted movements.

## **3. Methodology**

The serial correlation test and support vector machines classification model are used to test the weak form of EMH in this study.

### *3.1 Serial Correlation Test*

The serial correlation test is most frequently used for studying patterns in share prices and in market returns (Bundoo, 2000). Correlation among price levels of a specific Stock over a certain period are determined using the following formula:

$$R_t = b_0 + b_1R_{t-1} + b_2R_{t-2} + \dots + b_iR_{t-i} + U \tag{1}$$

Where:  $R_t$  is the natural log return on a particular Stock at time  $t$  being  $\ln(R_t/R_{t-1})$

$B_0$  is the intercept term

$B_i$  is the slope coefficient of the lagged return

$R_{t-i}$  is the lagged return on the same Stock for the  $i^{\text{th}}$  period

$U$  is the error term

In case of an efficient market in the weak form, the slope coefficients are expected to be zero. Else, if any of the independent variables are found to be significant in explaining the market returns at time  $t$ , it would mean that the past stock prices are relevant in determining future returns. If the null hypothesis of no serial correlation between current returns and lagged returns of a stock is rejected for the stocks in the sample studied, the market can be said to be inefficient in the weak form. As stated above, the input variables for the serial correlation test are lagged returns of the stock being examined. In this study, daily returns for sample stocks on the SEM for the years 2007 to 2016 are used to find any correlation between past and future returns.

### 3.2 SVM

According to the Vapnik (1999) model, SVM learns how to classify an object or observation into two or more classes (for example good or bad / increasing or decreasing) on a hyper-plane. To achieve the classification where the set of observations are not linearly separable, SVM transforms non-linear input into linear mode in a high dimension feature space using a non-linear function (also known as a kernel). It is an algorithm, which can learn classification and regression rules from data. In its simplest form, SVM attempts to generate a hyper-plane that separates two classes with the largest margin in between as per the Figure 2 below. The margin separating the nearest observations 1, 2, 3 and 4 is maximised as represented by the distance  $M$ , using the line  $Z3$  and the boundaries  $B1$  and  $B2$ . Vectors 1, 2, 3 and 4 are the support vectors. The area of the margin is called the hyper-plane. Any points predicted to be outside the margin, but which is found to be in the incorrect class or actually inside the margin is an error which the model then uses to re-adapt the hyper-plane.

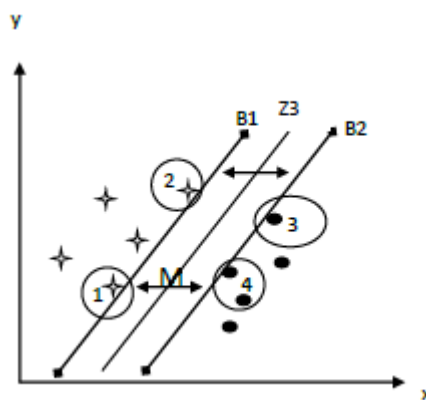


Figure 2. Hyper-plane constructed by SVM

The observations in the plot can be considered as a set of input data points:

$$D = \{(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i)\} \tag{2}$$

with each point having a binary classification  $y_j$  of either -1 or 1.

The SVM classification function takes the form:

$$F(x) = w \cdot x + b \tag{3}$$

To determine the class to which the point  $x_i$  belongs to, the SVMs require maximizing the distance between B1 and B2, or minimizing the inverse being:

$$\text{Min } \|w\|/2 \tag{4}$$

The minimization function is a quadratic optimisation problem which can be solved using Lagrange multipliers following which the classification equation is transformed into:

$$F(\mathbf{x}) = \sum_i \alpha_i y_i \mathbf{x}_i \cdot \mathbf{x} - b \tag{5}$$

To achieve the classification where the set of observations are not linearly separable, SVM transforms non-linear input into linear mode in a high dimension feature space using a non-linear function (also known as a kernel). Pan (2014) advocates the use of the Radial Basis Function (RBF) kernel:  $K(x,y) = \exp(-\gamma\|x-y\|^2)^d$  over the two other popular ones called the polynomial kernel and the sigmoid kernel for the following reasons: RBF involves fewer numerical difficulties and can handle the nonlinear models, the polynomial kernel has more hyper-parameters than the RBF kernel and the sigmoid kernel behaves like RBF kernel for certain parameters.

The minimisation problem can be fitted with a cost function such that it is transformed into:

$$C \sum_{i=0}^l (\xi_i + \xi_i^*) + \frac{1}{2} \|w\|^2$$

Where:

C is a regularisation parameter with small values allowing constraints to be easily ignored and large values making constraints hard to ignore

$\xi_i$  are points which are incorrectly classified by the model.

Model inputs for the SVM are as observed in literature (Kim, 2003; Ni et al., 2011; Rosillo et al., 2013 and Rosillo et al., 2014) and are as described in table 1 below. These inputs are based on strategies adopted by market traders and which try to depict trends in the performance of a stock. Where the notations are similar among the formulas, they are described only once.

Table 1. Input variables for the support vector machine model

Techniques	Description	Formula
%K	Stochastic %K. It compares where a security's price closed relative to its price range over a given time period.	$\frac{C_t - LL_{t-n}}{HH_{t-n} - LL_{t-n}} \times 100,$ <p>where <math>LL_t</math> and <math>HH_t</math> mean the lowest low and highest high of the Stock price in the last t days, respectively. <math>C_t</math> is the Stock price at time t.</p>
%D	Stochastic %D. It is the moving average of %K.	$\frac{\sum_{i=0}^{n-1} \%K_{t-i}}{n}$ <p>A 3 day moving average in the study.</p>

Slow %D	Stochastic slow %D. It is the moving average of %D.	$\frac{\sum_{i=0}^{n-1} \%D_{t-i}}{n}$ <p>A 3 day moving average.</p>
EMA	Exponential Moving Average. It is similar to a simple moving average, except than more weight is given to the latest prices.	$EMA_n(i) = \alpha \times S_i + (1 - \alpha) \times EMA_n(i - 1)$ $\alpha = \frac{2}{1+n}$ <p>Where: <math>S_t</math> is the current Stock price and</p> <p>A 5 day average is used in this study.</p>
Momentum	It measures the amount that a security's price has changed over a given time span.	$C_t - C_{t-4}$ <p>Where <math>C_{t-4}</math> is the 4 day lagged Stock price.</p>
ROC	Price rate-of-change. It displays the difference between the current price and the price n days ago.	$\frac{C_t}{C_{t-n}} \times 100$
Williams' %R	Larry William's % R. It is a momentum indicator that measures overbought /oversold levels.	$\frac{H_n - C_t}{H_n - L_n} \times 100$ <p>Where <math>H_n</math> and <math>L_n</math> is the highest high and lowest low respectively at period n. The period used in this study is 14 days.</p>
A/D Oscillator	Accumulation/distribution oscillator. It is a momentum indicator that associates changes in price.	$\frac{H_t - C_{t-1}}{H_t - L_t}$
Disparity5	5-day disparity. It means the distance of current price and the moving average of 5 days.	$\frac{C_t}{MA_5} \times 100$ <p>Where <math>MA_5</math> is the moving average of the Stock price over the last 5 days, including the current day.</p>
Disparity10	10-day disparity.	$\frac{C_t}{MA_{10}} \times 100$ <p>Where <math>MA_{10}</math> is the moving average of the Stock price over the last 10 days, including the current day.</p>

OSCP	Price oscillator. It displays the difference between two moving averages of a security's price.	$\frac{MA_5 - MA_{10}}{MA_5}$
CCI	Commodity channel index. It measures the variation of a security's price from its statistical mean.	$\frac{(M_t - SM_t)}{(0.015 D_t)}$ <p>Where <math>M_t = (H_t + L_t + C_t)/3</math></p> $SM_t = \frac{\sum_{i=1}^n M_{t-i+1}}{n} \quad D_t = \frac{\sum_{i=1}^n  \dot{M}_{t-i+1} - SM_t }{n}$ <p>Where <math>SM_t</math> is the mean deviation over the t time period. For the purpose of this study, 20 day period is used.</p>
RSI	Relative strength index. It is a price following an oscillator that ranges from 0 to 100.	$100 - \frac{100}{1 + (\sum_{i=0}^{n-1} U p_{t-i}/n) / (\sum_{i=0}^{n-1} D w_{t-i}/n)}$ <p>Where: <math>U_{pt}</math> means upward-price-change and <math>D_{wt}</math> means downward-price-change at time t. A 14 days period is used in this study.</p>
MACD	Moving Average Convergence Divergence. Designed mainly to identify trend changes.	$MACD(n) = EMA_k(i) - EMA_d(i)$ <p>Where k and d are different time periods. A 12 days and 26 days period is used in this study, for k and d respectively.</p>

Some of the inputs, such as %K, A/D Oscillator and CCI use the high and low prices of a Stock on a particular day, in addition to the closing price. Given the dataset obtained from the SEM did not include the high and low observations, the closing price on the day for which these data are required is used as proxy in the equations.

Being a classifier model, there are no regressions involved. The model is used to try to predict movements in prices of a Stock, which is thus measured in ordinal scale. In this respect, the classes applicable are as per table 2.

Table 2. Classification of movement in prices for prediction by SVM

Class Name	Description
Up	The price on the next day increased
Hold	The price on the next day remained the same
Down	The price on the next day decreased

Since the above comprise of 3 classes, it is not stated as a binary classification problem. In this case, SVM uses the “one versus the rest” method to construct a set of binary classifiers, each one trained to separate one class from the two others and combine the results to get a multi-class output.

To determine the accuracy of the model, the sample used is split into a training set comprising of 75% of the sample size to obtain the support vectors and a test set with the remaining 25% sample to predict the direction of Stock prices by using the results of the training set. The observations in the training set and test set are as allocated randomly from

the sample rather than the allocation being split in time brackets. We then look at the confusion matrix where predicted results are compared to actual results to determine the accuracy of the prediction.

For arriving at the results, the `e1071` library package in the R-Studio statistical software is used. To obtain the optimum cost function for use in the model, the `tune` function from the library which performs the classification using a range of costs to determine the best margin is used and for which a cost of 100,000 is identified.

### 3.3 Data Sources and Sampling

To test the weak form of market efficiency, the explanatory power of previous stock price information on price levels or whether returns on Stocks can be predicted using past stock prices is analysed. For the purpose of the above, time series data on stock prices is obtained from SEM and its fact books. Daily Stock prices from November 2006 to 31 January 2017 are collected for calculating the model inputs required for the SVM. A start period of November 2006 and end period of 31 January 2017 are used given some model inputs for the SVM as at January 2007 requires lagged observations from previous periods and prediction of movement in prices are made for periods of more than 1 day. However, the explanatory variables computed for training the models starts as from January 2007 and ends at December 2016. The serial correlation test is also performed on daily stock prices for the period from January 2007 to December 2016, representing 2,467 observations.

As at April 2017, there were 43 companies listed on SEM. Since listed firms vary in size and a size effect has been identified in literature, stocks of different sizes are captured. The process of selecting stocks listed on the SEM is thus as follows: The companies listed on SEM were classified based on common industries in which they operate. It is to be noted that companies which were not listed during the year 2007 or no longer on SEM in the year 2016 are excluded from the sample. For each industry, the highest mean and lowest market capitalisation were computed, and the selected stocks are the ones with market capitalisation equal to or nearest to the mean for the corresponding industry. The selection process led to 17 companies to be included in the sample. Since the Sugar and Transport sectors included only 1 company each, listed prior to 2007, they were selected without the need to compute the different market capitalisations.

## 4. Results

### 4.1 Serial Correlation Test

The serial correlation test is performed using lagged returns over 8 day periods as explanatory variables and current daily returns as the dependent variable for the period January 2007 to December 2016. It would accordingly be interesting to check the properties of such returns during the period. As per Bundoo (2000), if Stock prices follow the random-walk model, the coefficients in the equation being tested are expected to be zero or close to zero and not statistically significant. This study follows proportion of movement of prices for each Stock being studied over the 10-year period. Throughout the sample, events of no movement constitute a majority. In 6 cases, namely for the Stocks ASL, CMPL, MEI, PBL, PIM and VIVO, the proportion of cases with no movements in share price versus otherwise is over 70%, with CMPL and PIM reporting at over 90%. For these Stocks, even if the relationship between current returns and lagged returns is found to be high, caution should be taken in interpreting the results as these cases present very little volatility. Stocks which present proportions of volatility of close to or more than 50% of the observations are AML, MCB, NMH, SBM and SUN. The diagnostics testing for presence of heteroscedasticity and normality of residuals as well as VIF test to detect presence of multicollinearity have been carried out. The *p-value* generated by the Breusch-Pagan test reported for MCB, PBL, PIM, UBP and VIVO are more than 0.5, hence indicating homoscedasticity. Since heteroscedasticity present a risk for estimation of the standard errors, as they affect the precision of the *p-value*, the White standard errors estimator procedure to obtain robust estimators is used for the other Stocks. It is also observed that the distribution is moderately skewed for most of the variables except for CMPL, ENL, NIT, PBL, PIM and VIVO while the Kurtosis is far from the benchmark number of 3, except for NIT. As per Fowdar et al. (2007), having a high value of kurtosis shows a leptokurtic distribution, from which it can be inferred that Stock return series deviates from the prior condition of random walk model that is returns are normally distributed. Upon analysis of the maximum Cook's D value obtained for the observations, it is seen that only CMPL, ENL and FINC have influential outliers and hence may need to be further examined. For this study, given heteroscedasticity is also found for these observations, the White standard errors estimates are relied on. The VIF values for all the Stocks approximately 1 and hence indicates that there are no multicollinearity issues.

The results of the OLS regression is presented in Table 3. Serial correlation is found in the first order for most of the Stocks, where the *t statistic* is found to be significant. The coefficients, for these cases, also present a positive relationship with current returns. Varying orders of serial correlation can also be observed for the different Stocks, with up to five orders for MCB. It is noted, however, that the coefficients for all independent variables for OMN are very close to zero relative to other Stocks. Overall and through extrapolation from the results of the sample in this study,



there seems to be evidence that a majority of Stocks on the SEM do not follow a Random-Walk. However, Fowdar et al. (2007) argue that serial dependence of the values may suggest slow adjustment to new information on the market, the presence of insider information or the lack of liquidity due to infrequent trading or movement of prices. The findings from Table 5, tend to indicate the lack of liquidity as the culprit for some of the Stocks. Given the evidence obtained, it can be inferred that there is existence of serial correlation among lagged returns and hence the SEM is not strictly weak-form efficient.

4.2 SVM

The SVM classifier model is also used to try to determine whether information from trends and as obtained from technical analysis on a Stock can be used to determine the direction of that Stock. The factors as per Table 1 are used for the purpose of training the model to obtain the support vectors and determine the optimal hyper-plane. The results are then used to predict the direction of Stock prices on the sample. The random number generator is the R-Studio software is used. The training set for each Stock varies around 1,858 and the training set around 609 observations (it varies slightly each time the split is done), with the total always remaining at 2,467 observations.

Since SVM is computationally heavy, the SVM is tested on 2 random Stocks for prediction of price movements for different time periods of 1 day forward, 5 days forward, 10 days forward and 20 days forward as a benchmark, to see if there are any improvements in predictions. Being a classifier model for which the support vectors are critical for trying to predict Stock prices, the number of cases in the different categories being looked at should be sufficient enough for the model to correctly depict the discriminatory margin. Table 4 shows that in almost all the Stocks, the number of cases where there were no movements is dominant. This can lead to the model being trained to better identify such situations of no activity rather than the more interesting cases of upward and downward changes.

Table 4. Momentum in Stock prices for FINC and NMH for 1 day forward

	<b>FINC</b>	<b>NMH</b>
UP	394	666
DOWN	457	781
HOLD	1,616	1,020

*Source: Computed by the authors*

FINC and NMH are randomly chosen and the prediction results through SVM classification are as demonstrated in Tables 5 and 6.

Table 5. SVM prediction on next day movement in share price for FINC

		<b>FINC</b>		
		<b>Actual observations 1 day forward</b>		
		<b>Down</b>	<b>Hold</b>	<b>Up</b>
<b>Predicted observations</b>	<b>Down</b>	20	49	18
	<b>Hold</b>	88	298	62
	<b>Up</b>	8	49	18
	<b>% Accuracy</b>	17.24	75.25	18.37

*Source: Computed by the authors*

Table 6. SVM prediction on next day movement in share price for NMH

		<b>NMH</b>		
		<b>Actual observations 1 day forward</b>		
		<b>Down</b>	<b>Hold</b>	<b>Up</b>
<b>Predicted observations</b>	<b>Down</b>	9	16	7
	<b>Hold</b>	182	221	150
	<b>Up</b>	5	9	2
	<b>% Accuracy</b>	4.59	89.84	1.26

*Source: Computed by the authors*

As expected, the model shows high accuracy for circumstances where the prices of the Stocks under study do not move but fail to predict the upward and downward trends. Based on a 5 days, 10 days and 20 days forward in future returns, the counted movements in market prices for FINC and NMH from the total of 2,467 observations for the sample are shown as per Table 7.

Table 7. Count of momentum of Stock prices over different forward periods for FINC and NMH

	Movements in market prices					
	5 days forward		10 days forward		20 days forward	
	FINC	NMH	FINC	NMH	FINC	NMH
<b>Up</b>	713	952	930	1007	1010	1016
<b>Down</b>	672	1,225	1,226	1,266	1,297	1,329
<b>Hold</b>	1,082	290	311	194	160	122
<b>Total</b>	2,467	2,467	2,467	2,467	2,467	2,467

*Source: Computed by the authors*

The number of observations in the Up and Down price movement classes increase with more forward periods for FINC, while observations under the Hold class for that Stock decrease dramatically as a consequence. For NMH, there is relatively little change. Tables 8 and 9 give the results for the 3 periods for FINC and NMH. For FINC, as the gap in future returns increases, the prediction accuracy for the Down class increases dramatically. For the Hold class for FINC, it decreases on a large scale as well. The accuracy for the Up class varies, with the highest accuracy being for the 10 days forward. For NMH, the accuracy of prediction for the Down class sharply increases when the period is changed to a 5 days forward returns and then stabilises to around 93%. The inverse is noted for the Hold class but with nil accuracy for the 5 days and 20 days forward returns. A marked improvement is noted in the Up class. Overall, the period with the best results seems to be the 10 day forward returns. However, the accuracy in prediction for the Hold and Up classes is found to be still very low.

An investigation is thus made on the input variables with respect to their usefulness in training the model. Table 10 shows the average levels of accuracy for each factor used individually under the SVM for the 10 days forward returns. The average is measured only for the Up and Hold classes as the Down class represent an accuracy ranging from 55% to 100% in the list and, if included, can weigh down the results too much.

Table 8. SVM prediction results for 5, 10 and 20 days forward for FINC

		FINC								
		Actual observations 5 days forward			Actual observations 10 days forward			Actual observations 20 days forward		
		Down	Hold	Up	Down	Hold	Up	Down	Hold	Up
<b>Predicted observations</b>	<b>Down</b>	43	15	5	275	59	170	276	33	210
	<b>Hold</b>	114	219	219	13	2	5	7	3	3
	<b>Up</b>	14	27	41	31	9	46	33	5	39
	<b>% Accuracy</b>	25.15	83.91	15.47	86.21	2.86	20.81	87.34	7.32	15.48

*Source: Computed by the authors*

Table 9. SVM prediction results for 5, 10 and 20 days forward for NMH

		NMH								
		Actual observations 5 days forward			Actual observations 10 days forward			Actual observations 20 days forward		
		Down	Hold	Up	Down	Hold	Up	Down	Hold	Up
<b>Predicted observations</b>	<b>Down</b>	301	71	185	288	39	220	324	22	210
	<b>Hold</b>	2	0	0	0	1	1	1	0	1
	<b>Up</b>	19	7	15	24	4	24	20	4	19
	<b>% Accuracy</b>	93.48	0.00	7.50	92.31	2.27	9.80	93.91	0.00	8.26

*Source: Computed by the authors*

While the score is below 30% for all the factors for Up and Hold classes, EMA and CCI score relatively higher than all the other factors. RSI also score higher than 20% for NMH. %K, %D and Slow %D, William's R and MACD score average. The other factors have a score being comparatively lower. The results for the Down class is still high, with the lowest scorers being CCI, EMA and RSI (with respect to NMH). The marginal contribution of the high scorers in at least one of the Stocks are summarised in Table 11. Overall, the combination EMA and RSI indicators seem to be the best performer. However, it is to be noted that on the individual scale, there does not seem to be a clear cut winner. For example, the combination CCI, EMA and RSI seems to better at predicting downward movements. For FINC, EMA and RSI better predicts no movement while the best predictors for that class for NMH are CCI and EMA.

Table 10. Average accuracy of SVM prediction results for FINC and NMH

Factor Loadings	Average accuracy % (Up and Hold classes)		Accuracy % (Down classes)	
	FINC	NMH	FINC	NMH
%K	17.34	9.06	90.91	88.12
%D	16.65	17.10	79.00	73.93
Slow %D	15.25	15.11	78.06	65.35
EMA	26.21	25.38	63.01	67.99
Momentum	5.93	6.63	91.85	86.32
ROC	3.85	5.44	93.10	91.75
Williams' %R	17.34	9.06	90.91	88.12
A/D Oscillator	0.00	0.00	100.00	100.00
Disparity5	0.68	0.00	99.69	100.00
Disparity10	0.90	0.00	99.06	99.67
OSCP	1.13	0.40	98.43	98.02
CCI	26.61	26.77	57.37	54.13
RSI	18.87	22.38	82.45	65.02
MACD	7.24	14.11	91.54	87.13

*Source: Computed by the authors*

The combination of factors EMA and RSI is applied to the full sample for which the results are presented in Table 12. It can be seen that in the majority of Down cases and in all of the Up cases, there is predictability of movement in Stock prices. The accuracy for no movement is less than 50%. However, overall, the number of instances of predictability being possible (i.e. more than 50%, 39 instance) is higher than failure of predictability (12 instances). Accordingly, it can be said that prediction of changes in direction of market prices is possible and that arbitrage opportunities exists, thereby leading to the conclusion that the SEM is not weak-form efficient.

Table 11. Marginal contribution of factors in SVM prediction

Factor Loadings	FINC (% Accuracy)			NMH (% Accuracy)		
	Down	Hold	Up	Down	Hold	Up
CCI + EMA + RSI	81.06	10.67	27.47	92.18	6.67	13.25
CCI + EMA	57.37	14.49	42.99	66.12	13.33	27.31
CCI + RSI	35.11	11.59	58.37	66.45	4.44	31.33
EMA + RSI	70.22	39.13	65.61	66.45	8.89	59.44

*Source: Computed by the authors*

Table 12. SVM prediction results for companies in sample using the EMA and RSI factors

Stocks	Accuracy %		
	Down	Hold	Up
AML	58.93	28.57	56.47
ASL	68.16	63.29	76.85
CMPL	70.59	78.99	61.76
ENL	69.11	49.52	76.57
FINC	70.22	39.13	65.61
IBL	55.04	29.49	71.91
MCB	45.97	19.15	75.71
MEI	65.17	76.76	64.71
NIT	69.23	59.78	71.67
NMH	70.22	39.13	65.61
OMN	74.48	30.59	62.40
PBL	70.31	69.43	79.77
PIM	45.04	85.89	56.69
SBM	62.02	19.23	65.40
SUN	60.59	29.73	79.37
UBP	64.89	45.16	82.76
VIVO	69.48	56.79	81.75
Accuracy (No. of instances)			
Accuracy < 50%	2	10	0
Accuracy > 50%	15	7	17
<i>Of which:</i>			
50% < Accuracy < 60%	2	2	2
60% < Accuracy < 70%	8	2	6
Accuracy > 70%	5	3	9
<i>Source: Computed by the authors</i>			

## 5. Conclusion and Implications of the Study

The study focuses on whether there is evidence that SEM is market efficient. The weak form of efficiency is studied for a sample of 17 Stocks over the period January 2007 to December 2016 using the serial correlation test with the perspective of whether current Stock returns was correlated with past events. A popular method of machine learning technique, the SVM was applied, to determine whether future Stock returns could be predicted using information gained from trend and technical analysis of past information as obtained from 14 techniques.

The results demonstrated that there was evidence of the SEM not being weak-form efficient. From the initial serial correlation test, most Stocks in the sample appeared to be correlated with past returns in the first order. Some Stocks also demonstrated a significant relationship with returns for more than a one-day lag. Furthermore, the SVM had a success rate of more than 50% in most cases in predicting returns, with a peak of between 60% and 70% for predicting upward and downward movement of Stocks, which contributes to evidence of the market not being efficient in the weak-form.

The findings of this study have certain implications, namely with respect to the following:

- (1) Investors have the prospect of adopting speculative strategies to earn abnormal profits successfully. In an event of an under-reaction or over-reaction by market participants, given they may not have homogeneous expectations from analysis of market trends as is often the portrayed trait of noise traders, these strategies can have adverse effects on the market;
- (2) Profits from trading based on information and advanced techniques and models are possible. Automated trading systems engineered to scan market sentiment, analyse new information and predict movements in Stock prices, which is becoming a popular form of sophisticated trading on larger stock markets, have merit for implementation on SEM on a purely technical basis. However, the level and frequency of profits given the low volatility prevailing on SEM may not justify the costs of implementing such systems;
- (3) The lack of efficiency on the market can lead to incorrect interpretation of the perceived value of listed companies, also often used as a benchmark for determining the value of similar unlisted companies, and inadequate formulation of policies by regulators.

The serial correlation test and SVM model use daily returns for analysis. However, given the lack of liquidity and trading volume on SEM along with considering transaction costs, further analysis on larger holding periods is warranted to corroborate the findings in line with trading strategies that may be adopted by investors.

Furthermore, different Stocks seemed to demonstrate correlation with different lagged period returns and SVM seemed to indicate that different technical indicators may have different accuracy levels both on the class being considered and for the Stock being studied. These prompt for further studies required on the specifics of the different sensitivities and the underlying factors such as corporate announcements, industry shocks, market sentiment, anomalies, etc... The SVM model is particularly popular for different types of analysis such as prediction based on investor sentiment from social media are being complemented with other types of machine learning techniques, giving scope for further research to improve the accuracy of the results. Comparative models in machine learning techniques have also not been tested on whether they produce better results than SVM in the Mauritius context. Finally, as indicated by Kim & Shamsuddin (2008), the return predictability found in this study may be subject to fluctuation overtime.

This study contributes to literature on the study of the Efficient Market Hypothesis on the Stock Exchange in Mauritius by considering more modern strategies employed by market traders and techniques compared to literature. It paved the way to demonstrating the use of machine learning techniques for determining predictability of stock returns on the local exchange. Future research could entail a study of impact of using kernels other than the Radial Basis Function, with respect to the Support Vector Machine technique, on performance of prediction. The study of the influence of shorter time periods for the training set as well different combinations of trading strategies specific to each stock rather than a standardised function may also be insightful.

## References

- Anghel, G. D. I. (2015). Stock market efficiency and the MACD. Evidence from countries around the world. *Procedia Economics and Finance*, 32, 1414-1431. [https://doi.org/10.1016/S2212-5671\(15\)01518-X](https://doi.org/10.1016/S2212-5671(15)01518-X)
- Bundoo, S. K. (2000). The Mauritius Stock Exchange: An Assessment. *Research Journal*, 3, University of Mauritius.
- Bundoo, S. K. (2000). The Mauritius Stock Exchange: Sectoral Analysis, Risk And Return. *Research Journal*, 3, University of Mauritius.
- Cheeneebas, J., Galanis, G., Gopaul, A., & Bhuruth, M. (2008). Improving Artificial Neural Network Forecasts with Kalman Filterin. *University of Mauritius Research Journal*, 14.
- Fama, E. F. (1965). The Behaviour of Stock-Market Prices. *Journal of Business*, 38(1), 34-105. <https://doi.org/10.1086/294743>
- Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 25(2). <https://doi.org/10.2307/2325486>
- Fowdar, S., Subadar, U. A., Lamport, M, Sannassee, R. V., & Fawzee, M. (2007) Assessing the Level of Efficiency of The Stock Exchange of Mauritius. *Research Journal, Special Issue*, 13A, University of Mauritius.
- Kara, Y., Boyacioglu, M. A., & Baykan, Ö. K. (2011) .Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. *Expert Systems with Applications*, 38, 5311-5319. <https://doi.org/10.1016/j.eswa.2010.10.027>
- Kim, J. H., & Shamsuddin, A. (2008). Are Asian stock markets efficient? Evidence from new multiple variance ratio tests. *Journal of Empirical Finance*, 15, 518-532. <https://doi.org/10.1016/j.jempfin.2007.07.001>
- Kim, K. J. (2003). Financial time series forecasting using support vector machines. *Neurocomputing*, 55, 307-319. [https://doi.org/10.1016/S0925-2312\(03\)00372-2](https://doi.org/10.1016/S0925-2312(03)00372-2)
- Ni, L. P., Ni, Z. W., & Gao, Y. Z. (2011) Stock trend prediction based on fractal feature selection and support vector machine. *Expert Systems with Application*, 38, 5569-5576. <https://doi.org/10.1016/j.eswa.2010.10.079>
- Pan, L. (2014). Prediction of the Moving Direction of Google Inc. Stock Price Using Support Vector Classification and Regression. *Asian Journal of Finance & Accounting*, 6(1). <https://doi.org/10.5296/ajfa.v6i1.5485>
- Patil, S. S., Patidar, K., & Jain, M. (2016). Stock Market Prediction Using Support Vector Machine. *International Journal of Current Trends in Engineering & Technology*, 2(1).
- Rosillo, R., Giner, J., & De la Fuente, D. (2014). Stock Market simulation using support vector machines. *Journal of Forecasting*, 33(6), 488-500. <https://doi.org/10.1002/for.2302>
- Rosillo, R., Giner, J., Puente, J., & Ponte, B. (2013) 'Different Stock Market Models Using Support Vector Machines. *International Journal of Trade, Economics and Finance*, 4(5). <https://doi.org/10.7763/IJTEF.2013.V4.307>
- Shen, S., Jiang, H., & Zhang, T. (2012). Stock Market Forecasting Using Machine Learning Algorithms', from: <http://cs229.stanford.edu/proj2012/ShenJiangZhang-StockMarketForecastingusingMachineLearningAlgorithms.pdf>
- Sheta, A., Ahmed, S., & Faris, H. (2015). A Comparison between Regression, Artificial Neural Networks and Support

Vector Machines for Predicting Stock Market Index. *International Journal of Advanced Research in Artificial Intelligence*, 4(7). <https://doi.org/10.14569/IJARAI.2015.040710>

Vapnik, V. (1999). An Overview of Statistical Learning Theory. *IEEE Transactions on Neural Networks*, 10(5). <https://doi.org/10.1109/72.788640>

Wang, Y. (2014). Stock price direction prediction by directly using prices data: an empirical study on the KOSPI and HIS. *International Journal of Business Intelligence and Data Mining*, 9(2), 145-160. <https://doi.org/10.1504/IJBIDM.2014.065091>

**Appendix**

Table 3. OLS regression results for serial correlation test

<b>MODEL USED: <math>R_t = b_0 + b_1R_{t-1} + b_2R_{t-2} + b_3R_{t-3} + b_4R_{t-4} + b_5R_{t-5} + b_6R_{t-6} + b_7R_{t-7} + b_8R_{t-8}</math></b>									
	b0	b1	b2	b3	b4	b5	b6	b7	b8
	[t statistic]	[t statistic]	(t statistic)	[t statistic]	[t statistic]	[t statistic]	[t statistic]	[t statistic]	[t statistic]
	(prob)	(prob)	(prob)	(prob)	(prob)	(prob)	(prob)	(prob)	(prob)
AML	-0.0001383 [-0.3136] (0.7538)	0.05570162 [1.0526] (0.2926)	-0.04488496 [-1.0541] (0.2920)	-0.02585013 [-0.7697] (0.4416)	-0.00447832 [-0.1580] (0.8745)	-0.01506619 [-0.4444] (0.6568)	-0.02951400 [-1.2084] (0.2270)	-0.01178325 [-0.4563] (0.6482)	-0.02506958 [-0.7119] (0.4766)
ASL	0.000166 [0.4243] (0.671386)	0.0948698 a [2.6280] (0.008642)	0.0762021 c [1.7305] (0.083664)	-0.0043408 [-0.1638] (0.869908)	0.0339361 [0.8373] (0.402479)	0.0005087 [0.0239] (0.980902)	-0.0200549 [-0.4602] (0.645449)	-0.0205432 [-0.9820] (0.326216)	-0.0035948 [-0.1856] (0.852751)
CMPL	-0.0002824 [-0.7117] (0.4767)	0.0987554 [1.1897] (0.2343)	0.0926650 [1.5926] (0.1114)	0.0212387 [0.3566] (0.7214)	-0.0118082 [-0.2831] (0.7771)	0.0317629 [0.7659] (0.4438)	0.0154796 [0.3578] (0.7206)	-0.0008843 [-0.0489] (0.9610)	-0.0046861 [-0.3403] (0.7337)
ENL	0.0004840 [1.1967] (0.23155)	0.1407554 b [2.0708] (0.03848)	-0.0389080 [-0.6488] (0.51652)	-0.0190294 [-0.4034] (0.68671)	-0.0078324 [-0.1585] (0.87406)	-0.0059465 [-0.1314] (0.89544)	-0.064474 b [-1.9696] (0.04899)	-0.0538957 [-1.2909] (0.19685)	0.0235011 [0.6986] (0.48486)
FINC	-8.762e-05 [-0.2717]	7.973e-02 b [2.5182]	-2.964e-02 [-0.3829]	-2.866e-02 [-1.0669]	-7.268e-03 [-0.1852]	-4.323e-02 b [-2.0522]	9.534e-03 [0.2685]	1.519e-03 [0.0778]	2.402e-02 [0.9800]

	(0.78591)	(0.01186)	(0.70185)	(0.28612)	(0.85310)	(0.04026)	(0.78834)	(0.93801)	(0.32717)
IBL	0.0007274 b [2.3501] (0.01885)	0.0615406 [1.4154] (0.15707)	-0.0376698 [-1.0648] (0.28707)	-0.0215203 [-0.5441] (0.58643)	-0.0375721 [-1.4721] (0.14112)	-0.0310672 [-0.7585] (0.44820)	-0.0302547 [-1.3368] (0.18141)	-0.0372507 [-0.8544] (0.39294)	-0.0042629 [-0.1234] (0.90177)
MCB	0.0002925 [1.415] (0.15710)	0.1341776 a [6.657] (3.43e-11)	0.0109268 [0.538] (0.59075)	-0.0085146 [-0.419] (0.67499)	0.0528692 a [2.607] (0.00918)	-0.0521620 b [-2.572] (0.01016)	-0.0362067 c [-1.783] (0.07467)	0.0440189 b [2.167] (0.03035)	0.0227838 [1.131] (0.25834)
MEI	0.0001330 [0.4502] (0.6526)	0.1308253 a [2.5320] (0.0114)	-0.0362140 [-0.6505] (0.5154)	0.0124148 [0.3445] (0.7305)	-0.0176897 [-0.4526] (0.6509)	0.0111120 [0.6889] (0.4909)	-0.0277789 [-0.7299] (0.4655)	-0.0159570 [-0.3535] (0.7237)	-0.0155249 [-0.7107] (0.4774)
NIT	0.0004157 [1.1647] (0.24427)	0.0761316 c [1.8459] (0.06502)	0.0473571 [1.3139] (0.18901)	0.0269850 [0.7570] (0.44909)	-0.047517 c [-1.8946] (0.05826)	-0.0289043 [-1.3229] (0.18598)	0.0097430 [0.4495] (0.65309)	-0.0059197 [-0.2196] (0.82622)	-0.0312092 [-1.1336] (0.25707)
NMH	-0.0002017 [-0.5861] (0.557835)	0.1436948 a [2.6495] (0.008113)	-0.0051692 [-0.1183] (0.905854)	-0.0037298 [-0.0991] (0.921101)	0.0083660 [0.2625] (0.792933)	0.0421171 [1.3043] (0.192245)	-0.0057778 [-0.2244] (0.822439)	0.0207176 [0.7645] (0.444631)	0.0106731 [0.4262] (0.670012)
OMN	2.137e-05 [0.0795] (0.93661)	7.591e-02 [1.4917] (0.13591)	4.550e-02 [1.2526] (0.21048)	-6.071e-02 [-1.2475] (0.21235)	3.362e-02 [0.8459] (0.39771)	-1.518e-02 [-0.4246] (0.67120)	-3.395e-02 [-1.4520] (0.14664)	-6.043e-02 [-0.5013] (0.13341)	4.760e-02 c [1.7313] (0.08352)
PBL	0.0006351 a [2.803] (0.00510)	-0.0155881 [-0.773] (0.43966)	-0.0149239 [-0.740] (0.45941)	0.0565486 a [2.805] (0.00508)	0.0443331 b [2.196] (0.02815)	0.0317219 [1.572] (0.11616)	-0.0278926 [-1.383] (0.16666)	0.0018290 [0.091] (0.92774)	-0.0131717 [-0.653] (0.51370)
PIM	0.0001496 [0.892]	0.0305226 [1.513]	0.0375656 c [1.862]	-0.0047448 [-0.235]	0.0369856 c	0.0038313 [0.190]	0.0165275 [0.816]	0.0012105 [0.060]	0.0138673 [0.685]

	(0.3727)	(0.1303)	(0.0628)	(0.8142)	[1.832] (0.0671)	(0.8495)	(0.4148)	(0.9523)	(0.4934)
SBM	0.0001577 [0.5181] (0.6044)	-0.0191298 [-0.3431] (0.7316)	-0.0069162 [-0.1599] (0.8730)	0.0200633 [0.4926] (0.6224)	0.0018772 [0.0502] (0.9600)	-0.0283829 [-0.9017] (0.3673)	-0.0209192 [-0.6167] (0.5375)	0.0072157 [0.1937] (0.8464)	0.0489824 [1.3163] (0.1882)
SUN	-0.0001840 [-0.4895] (0.62456)	0.0928415 c [1.8602] (0.06298)	-0.0364583 [-0.7327] (0.46380)	-0.0374608 [-0.9058] (0.36511)	0.0144037 [0.4537] (0.65009)	0.0315250 [0.9025] (0.36688)	0.0173321 [0.5338] (0.59356)	0.0076937 [0.2534] (0.79996)	-0.0213792 [-0.8633] (0.38804)
UBP	0.0002785 [1.115] (0.265)	0.0787486 a [3.907] (9.58e-05)	0.0271219 [1.342] (0.180)	0.0195582 [0.968] (0.333)	0.0178812 [0.885] (0.376)	0.0120614 [0.597] (0.551)	0.0159909 [0.791] (0.429)	0.0124764 [0.617] (0.537)	-0.0174447 [-0.866] (0.387)
Vivo	0.0001936 [0.759] (0.4480)	0.1243753 a [6.166] (8.14e-10)	-0.0147133 [-0.724] (0.4691)	0.0386555 c [1.902] (0.0573)	0.0275501 [1.355] (0.1754)	-0.0131347 [-0.646] (0.5183)	0.0116664 [0.574] (0.5657)	0.0176126 [0.868] (0.3856)	-0.0055801 [-0.277] (0.7818)

Source: Computed by the authors

a: significant at <0.001 level, b: significant at 0.01 level, c: significant at 0.05 level, d: significant at 0.1 level.

### Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the [Creative Commons Attribution license](#) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.