Current Approaches

Fundamental Analysis:

- Fundamental analysis is the analysis of a businesss financial information or statements like financial figures, balance sheet, businesss assets, and other information publicly available.

Technical Analysis:

- Technical analysis, employs models for predicting the direction of market prices through the study of market historic data, is mostly used by market analyst.

Safdar, (2016), "Industry competition and fundamental analysis.Journal of Accounting Literature, Vol.37, pp.3654.

Nazrio et al.(2017), "A literature review of technical analysis on stock markets". The Quarterly Review of Economics and Finance.

Technical Analysis

- Market indicators add significant depth to technical analysis, because they contain much more information than price and volume. There are large number of technical indicators which are divided into two major categories, Leading and Lagging indicator.

Shynkevich et al., 2017, "Forecasting price movements using technical indicators: Investigating the impact of varying input window length". Neurocomputing, Vol.264 pp.7188, 2017

Bollinger Band [BB]

It is based on the volatility of given stock price and consists

- An N-period moving average (MA)
- Standard Deviation of N-period $= \sigma$
- An upper band (MA + K σ)
- An lower band (MA - K σ)
- The middle band lies between MA $+K\sigma$ and MA $-K\sigma$.

Bollinger, J.(2002). "Bollinger on Bollinger Bands" Published by McGraw-Hill,ISBN 0-07-127368-3,2002

Trading Band for analyze buy/ sell pattern

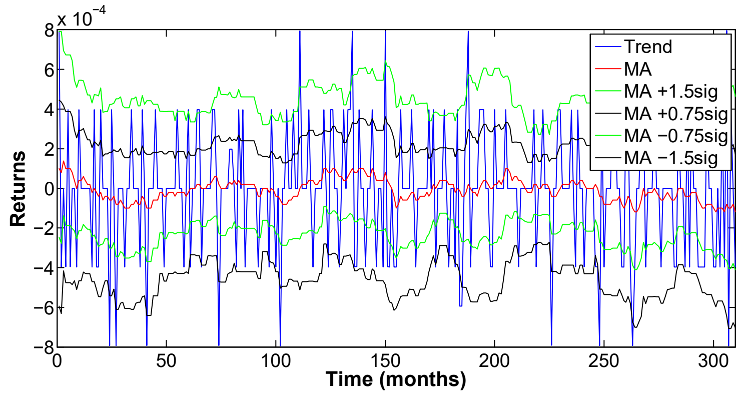

Figure: BB of Crude Oil minute data 22 Mar 2011

Optimal Band

Construction of Optimal Band

- Consider the time series returns $X_1, X_2..., X_n$
- $\alpha(AbsoluteMax) = Max(X_1, X_2, ..., X_n)$
- $\delta(AbsoluteMin) = Min(X_1, X_2, ..., X_n)$
- $\beta_i = Max(X_i, X_{i+1}, ..., X_{i+4}),\ 1 \leqslant i \leqslant n-4$
- $\gamma_i = Min(X_i, X_{i+1}, ..., X_{i+4}),\ 1 \leqslant i \leqslant n-4$
- $$f(\alpha, \beta, \gamma, \delta) = a * \alpha + b * \overline{\beta} + c * \overline{\gamma} + d * \delta$$

$\overline{\beta}$ *is mean of all* β_i, $\overline{\gamma}$ *is mean of all* γ_i

We now solve the following optimization problem to obtain the values of parameters (a,b,c,d)

$$\underbrace{Max}_{\text{a, b, c, d}} \; f(\alpha, \bar{\beta}, \bar{\gamma}, \delta) = a\alpha + b\bar{\beta} + c\bar{\gamma} + d\delta$$

s.t.

$$\begin{cases} f(\alpha, \bar{\beta}, \bar{\gamma}, \delta) \leqslant Min((\beta_i - \gamma_i)/2, (\alpha - \beta_i)), \\ 1 \leqslant i \leqslant n-4 \\ f(\alpha, \bar{\beta}, \bar{\gamma}, \delta) \geqslant 0 \end{cases}$$

We use Solver in MS excel to solve this optimization problem.

Continue...

- **Upper Band**$[UB_1] = \beta_i + f(\alpha, \beta, \gamma, \delta)$
- **Upper Band**$[UB_2] = \beta_i - f(\alpha, \beta, \gamma, \delta)$
- **Middle Band**$[MB] = f(\alpha, \beta, \gamma, \delta)$
- **Lower Band**$[LB_1] = \gamma_i + f(\alpha, \beta, \gamma, \delta)$
- **Lower Band**$[LB_2] = \gamma_i - f(\alpha, \beta, \gamma, \delta)$

Vijay, V. and Paul, P. (2015). A New Trading Band for Prediction of Buy and Sell Signals and Forecasting of States. *International Journal of Applied Management Sciences and Engineering*, 2(2), 34-54.

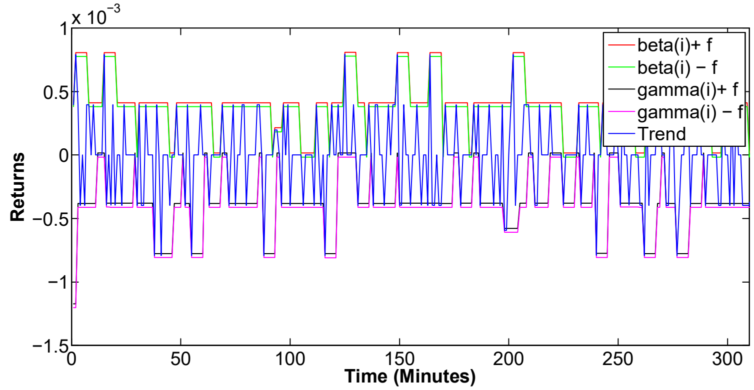

Figure: Optimal band of Crude Oil close price 22 Mar 2011

Advantage of OB Over the Other Technical Indicators

- **Decision Time**
  - Decision time for buy or sell in other model and Bollinger Band is small than that of Optimal Band.
  - In optimal band, decision time is slightly large due to its constant value for 2-3 time ticks.
- **Under High Volatile Conditions**
  - Other models follow same state for more time ticks.
  - The Bollinger bands are smooth, they do not completely follow the market.
  - The optimal band follows the market after 2-3 time ticks, if not immediately.

Hidden Markov Model for State Prediction

Markov Model: A sequence of discrete random variables S_t : t ϵ N is said to be a (discrete-time) Markov chain (MC) if for all t ϵ N it satisfies the Markov property and P known as transition probability.

$$P(S_{t+1} \mid S_t, ..., S_1) = P(S_{t+1} \mid S_t)$$

Hidden Markov Model: A hidden Markov model X_t : t ϵ N is a particular kind of dependent mixture. With X(t) and S(t) representing the histories from time 1 to time t, as follow:

$$P(S_t \mid S_{t-1}) = P(S_t \mid S_{t-1}), t = 2, 3, ...$$
$$P(X_t \mid X_{t-1}, S(t)) = P(X_t \mid S_t), t \epsilon N.$$

where X = X_1, X_2,...,X_t is data point at time t.

Three Basic Problems for HMM

- Evaluation: Given the model $\lambda = (A, B, \pi)$, how do we compute the probability of occurrence of the observation sequence O = o_1, o_2,...,o_t are corresponding to data point.
- Decoding: Given an observation sequence and a model λ, how do we choose a state sequence $S = \{S_1, S_2, ..., S_N\}$, that best explains the observations.
- Training: Given the observation sequence and a space of models found by varying the model parameters $\lambda = (A, B, \pi)$, how do we find the model that best explains the observed data.

Rabiner,L. R.(1989). A tutorial on hidden markov models and selected applications in speech recognition: In Proceedings of IEEE, Vol (77),257.

TPM of Crude Oil Minute Data 22 Mar 2011

Table: Relative Strength Index

	S1	S2	S3
S1	0.9386	0.0604	0.001
S2	0.0870	0.9016	0.0184
S3	0.0001	0.3332	0.6667

Table: Bollinger Band

	S1	S2	S3
S1	0.1256	0.7500	0.1250
S2	0.2800	0.5200	0.200
S3	0.3333	0.6667	0.0000

	S1	S2	S3
S1	0.2588	0.4647	0.2765
S2	0.4500	0.2444	0.3056
S3	0.3500	0.2000	0.4500

Table: Optimal Band

Probability of Predicted State Sequence

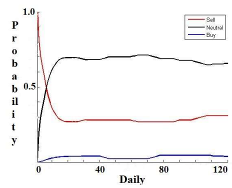

Figure: Relative Strength Index

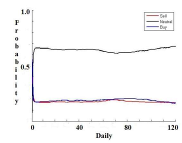

Figure: Bollinger Band

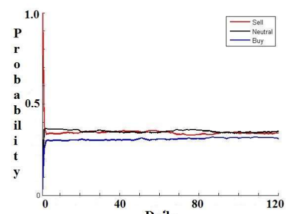

Figure: Optimal Band

Table: Comparison Table

Model	Total States	Match States	Error
RSI	120	89	25.84
Bollinger Band	120	91	24.17
Optimal Band	**120**	**93**	**22.36**

Paul, P., Vijay, V. (2015). Comparison of Buying and Selling Pattern of the Stocks Using Different Technical Indicators. September 2015, Conference: National Conference at Central University Rajasthan.

Hidden Markov Model

Now we extend Markov model to Hidden Markov Model by dividing each state into observation sequence is Sq_{i1}, Sq_{i2}, Sq_{i3}, Sq_{i4} and Sq_{i5}, $1 \leq i \leq 3$,. But observation is probabilistic function of state, defined as

$$Sq_{i1} = (0 \ to \ 20\% \ of \ S_i)$$

$$Sq_{i2} = (20 \ to \ 40\% \ of \ S_i)$$

$$Sq_{i3} = (40 \ to \ 60\% \ of \ S_i)$$

$$Sq_{i4} = (60 \ to \ 80\% \ of \ S_i)$$

$$Sq_{i5} = (80 \ to \ 100\% \ of \ S_i)$$

$$p(Sq_{ij}) = \frac{\text{No of data in } Sq_{ij}}{\text{No of data in state } S_i}$$

Observation Probability Matrix of Crude Oil for RSI and BB

	Sq1	Sq2	Sq3	Sq4	Sq5
S1	0.4000	0.1000	0.2000	0.0000	0.3000
S2	0.2593	0.3533	0.2852	0.0211	0.0743
S3	0.1000	0.1000	0.4000	0.0000	0.4000

	Sq1	Sq2	Sq3	Sq4	Sq5
S1	0.0492	0.0044	0.3000	0.001	0.6442
S2	0.0281	0.2000	0.7607	0.0044	0.0056
S3	0.0285	0.0283	0.0000	0.1000	0.8430

Table: Probability Distribution of observation sequence of OB

	Sq1	Sq2	Sq3	Sq4	Sq5
S1	0.3966	0.0172	0.5172	0.0000	0.0690
S2	0.0722	0.0000	0.8763	0.0100	0.0415
S3	0.0769	0.0171	0.5214	0.0110	0.3736

- Use function "hmmgenerate" randomly generates series of sequence and states
- The most likely states sequence generate using inbuilt MATLAB function
- $[likelystates] = hmmviterbi(sequence, prob - SB, prob - OS)$

State sequences and sub-sequences using MATLAB tools:

RSI

Seq $Sq_0 \rightarrow 2 \rightarrow 1 \rightarrow 1 \rightarrow 4 \rightarrow 1 \rightarrow 4 \rightarrow 2 \rightarrow 5 \rightarrow 1 \rightarrow 2$

States $S_0 \rightarrow 2 \rightarrow 3 \rightarrow 1 \rightarrow 2 \rightarrow 1 \rightarrow 2 \rightarrow 1 \rightarrow 2 \rightarrow 2 \rightarrow 2$

Likely-states $S_0 \rightarrow 2 \rightarrow 2 \rightarrow 1 \rightarrow 2 \rightarrow 1 \rightarrow 2 \rightarrow 2 \rightarrow 3 \rightarrow 1 \rightarrow 2$

BB

Seq $Sq_0 \rightarrow 5 \rightarrow 5 \rightarrow 5 \rightarrow 3 \rightarrow 3 \rightarrow 2 \rightarrow 5 \rightarrow 3 \rightarrow 3 \rightarrow 3$

States $S_0 \rightarrow 1 \rightarrow 3 \rightarrow 1 \rightarrow 2 \rightarrow 2 \rightarrow 3 \rightarrow 1 \rightarrow 2 \rightarrow 2 \rightarrow 2$

Likely-states $S_0 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 2 \rightarrow 2 \rightarrow 3 \rightarrow 1 \rightarrow 2 \rightarrow 2 \rightarrow 2$

OB

Seq $Sq_0 \rightarrow 1 \rightarrow 3 \rightarrow 5 \rightarrow 3 \rightarrow 1 \rightarrow 3 \rightarrow 5 \rightarrow 3 \rightarrow 3 \rightarrow 3$

States $S_0 \rightarrow 1 \rightarrow 1 \rightarrow 3 \rightarrow 3 \rightarrow 1 \rightarrow 2 \rightarrow 2 \rightarrow 2 \rightarrow 3 \rightarrow 3$

Likely-States $S_0 \rightarrow 1 \rightarrow 1 \rightarrow 3 \rightarrow 1 \rightarrow 1 \rightarrow 1 \rightarrow 3 \rightarrow 2 \rightarrow 2 \rightarrow 2$

Log Linear Model

- Log-linear analysis examine the relationship between more than two categorical variables
- Used for both hypothesis testing and model building
- Pearson's chi-square test only allows for two of the variables to be compared at a time
- But Log linear model is generalized model for more than two variables

Contingency Table

- It is essentially a display format used to analyse and record the relationship between two or more categorical variables
- It is the categorical equivalent of the scatter plot used to analyse the relationship between two continuous variables
- The term contingency table was first used by the statistician Karl Pearson in 1904
- We have used 2-Dim and 3-Dim contingency table using Pattern, Returns and Volume of transaction as variable.

Karl Pearson, F.R.S. (1904). Mathematical contributions to the theory of evolution. Dulau and Co.

Log Linear Model

A three way contingency table is defined as

$$l^{(3)}(i_1 i_2 i_3) = \tau^{(3)}_{123}(i_1 i_2 i_3) + \tau^{(3)}_{12}(i_1 i_2) + \tau^{(3)}_{13}(i_1 i_3)$$

$$+\tau^{(3)}_{23}(i_2 i_3) + \tau^{(3)}_{1}(i_1) + \tau^{(3)}_{2}(i_2) + \tau^{(3)}_{3}(i_3) + \tau^{(3)}_{\phi}$$

$$= \sum_{A} \tau^{(3)}_{A}(i_A)$$

A two way contingency table is defined as

$$l^{(2)}(i_1 i_2) = \tau^{(2)}_{12}(i_1 i_2)$$

$$+\tau^{(2)}_{1}(i_1) + \tau^{(2)}_{2}(i_2) + \tau^{(2)}_{\phi} = \sum_{Z \subseteq 1,2} \tau^{(2)}_{Z}(i_Z).$$

Vellaisamy, P., Vijay, V., 2007. Some collapsibility results for n-dimensional contingency tables. Ann. Inst. Statist.

Log Linear Hypothesis

- Within the log-linear parametrization a number of important hypothesis can be formulated by setting larger or smaller sets of τ to 0. The following are the main type.

H_1: $\tau_{ijk}^{ABC} = \tau_{ij}^{AC} = 0$
H_2: $\tau_{ijk}^{ABC} = \tau_{jk}^{AB} = 0$
H_3: $\tau_{ijk}^{ABC} = \tau_{ik}^{BC} = 0$
H_4: $\tau_{ijk}^{ABC} = \tau_{ij}^{AC} = \tau_{ik}^{BC} = 0$
H_5: $\tau_{ijk}^{ABC} = \tau_{jk}^{AB} = \tau_{ij}^{BC} = 0$
H_6: $\tau_{ijk}^{ABC} = \tau_{ik}^{AB} = \tau_{jk}^{AC} = 0$
H_7: $\tau_{ijk}^{ABC} = \tau_{ik}^{AB} = \tau_{jk}^{BC} = \tau_{jk}^{AC} = 0$

Anderson E.B. (1996). Introduction to the statistical analysis of categorical data. Springer.

Experiment and Result of 2-Dim Table

- Maruti Suzuki Co. 23 Nov 07 to 23 Nov 09 daily frequency

	H	M	L	Total
S	09	66	61	136
N	37	159	15	211
B	54	71	06	131
Total	90	296	82	478

- Hypothesis define in term of Z(H) value. Z(H) = 182.634 $\chi^2_{0.95} = 9.488$ at degree 4, Z(H) $\geq \chi^2_{0.95}$ Z(H) value so that hypothesis is rejected.

Table: 3-dim Frequency table of 23 Nov 07 to 23 Nov 09

		C: Volume	
A: Pattern	B: Returns	Up	Down
Sell	High	02	07
	Moderate	42	24
	Low	33	28
Neutral	High	17	20
	Moderate	78	81
	Low	05	10
Buy	High	28	26
	Moderate	44	27
	Low	04	02

Expriment and Result of 3-Dim Table

Table: Hypothesis and Z(H) value table of Maruti Suzuki 23 Nov 07 to 23 Nov 09

hypo	Z(H) Value	Df	p-value
H_1	11.312	6	0.07919
H_2	132.384	8	0.00001
H_3	8.843	6	0.33713
H_4	13.546	8	0.19472
H_5	134.618	10	0.00001
H_6	137.087	10	0.00001
H_7	139.321	12	0.000001

Vijay, V. and Paul, P. (2016). Analyzing Returns and Pattern of Financial Data Using Log- Linear Modeling. Computational Research, 4(1), 1-7.

Turn of the year effect

Definition: The Turn-of-the-Year effect refers to the phenomenon that small stocks have unusually high returns during the period beginning on the last trading day of december and continuing through january [Ritter (1988)].

- The average return for the month of January was 3.48 % percent compared to only 0.42 % per month for the other eleven months. The turns-of-the-Year effect occurs regularly in various stock markets
- Give new evidence with the help of logistic regression and buy/ sell ratio.

Ritter Jay R.(1988). "The Buying and Selling Behaviour of Individual Investor at the Turn of the Year." *JOF 43*.

Logistic Regression: In form of linear regression equation generates the predict probability coefficients b_0, b_1,...,b_l with different measurements X_1, X_2,...,X_n, it could be closed price, open price or difference of closed and open price etc.

$$\text{Reg}(month) = b_0 + b_1X_1 + b_2X_2 + ... + b_nX_l$$

$$\text{Reg}(month) = \ln(\frac{p}{1-p})$$

$$p(y = 1|x_1, ...x_n) = \frac{1}{1 + \exp -(b_0 + \sum_k b_k x_k)}.$$

where p is the probability of presence of the characteristic of interest.

Pampel, F. (2000). *"Logistic Regression: A primer."* Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-132. Thousand Oaks, CA: Sage Publications.

Table: Regression line of Tata Motor 2004-2007

	2005		2006	
	S=a+b*B	S=c+d*N	S=a+b*B	S=c+d*N
J	0.28219	0.56069	-0.081772	-0.66873
	214.86	38.552	200.78	155.13
F	2.8282	3.3017	-0.313	2.4854
	1711.6	1274.3	1576.5	501.1
M	1.0154	0.76169	-3.6383	0.77058
	390.32	141.59	880.53	373.76
A	2.0321	1.0901	0.59805	1.2132
	410.11	175.11	333.97	133.97
M	-5.3796	-1.0464	0.029227	-0.0055596
	820.09	306.18	152.52	70.442
J	-2.2403	-0.017829	1.9097	1.3738
	791.42	378.99	150.41	73.994
J	2.2425	2.5059	-0.33509	0.83319
	720.36	495.66	158.17	45.496
A	-0.6923	0.33061	-4.2042	-1.1456
	453.39	154.55	1029.7	391.02
S	-1.5979	-3.2231	-2.8569	1.8696
	440.42	345.48	880.99	406.35
O	0.99681	1.2202	0.051135	-0.64262
	354.33	118.8	256.11	12.455
N	-2.249	-3.8557	-2.9913	-0.76799
	324	279.63	1143.6	200.39
D	-0.82806	0.32374	0.21309	-1.6977
	397.43	154.71	108.68	174.14

The Probability of Significant Coefficient of Linear Regression Equation for Sell and Buy

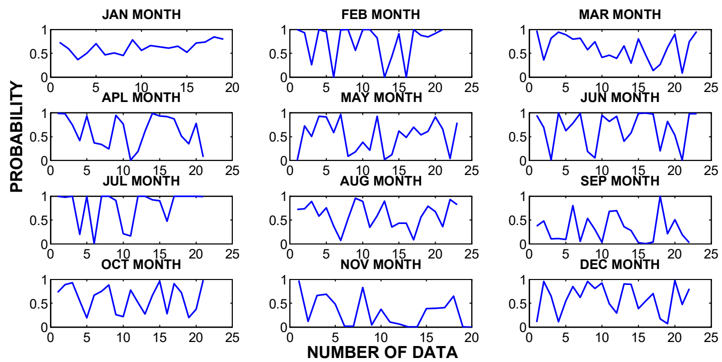

Figure: Tata Motor 2005

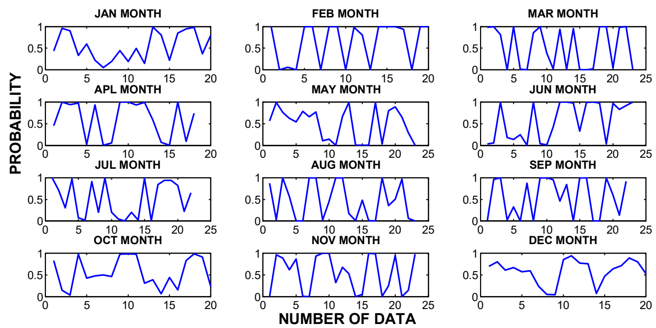

Figure: Tata Motor 2006

Definition of Buy/ Sell Ratio

Buy/ Sell Ratio: Buy sell ratio mean ratio of number of buying stocks to number of selling stocks. The buy and sell pattern of a stock are obtained from optimal band [Vijay and Paul (2015)]. Analysis of turn of the year effect or January effect based on buy/ sell ratio. It defined as large capital stocks has small return in January month.

Lakonishokand Josef Smidt Seymour (1984). "Volume and Turn of the Year Behaviour."Journal of Financial Economics 13, 435-56.

Formula to Find Buy/ Sell Ratio

Suppose total time ticks data of January month = N
Window size m = 6
Number of buy, sell and neutral pattern in single window

$$\sum B_i, ST_i \leq B_i \leq ST\,i + m - 1 \qquad \forall 1 \leq i \leq N - m$$

$$\sum S_i, ST_i \leq S_i \leq ST\,i + m - 1 \qquad \forall 1 \leq i \leq N - m$$

$$\sum N_i, ST_i \leq N_i \leq ST\,i + m - 1 \qquad \forall 1 \leq i \leq N - m$$

The value of buy/ sell ratio = $\frac{B_i}{S_i}$
The value of buy/ neutral ratio = $\frac{B_i}{N_i}$
The value of sell/ neutral ratio = $\frac{S_i}{N_i}$
The range of the these ratio lies between 1/6 to 6.

Buy/ Sell Ratio

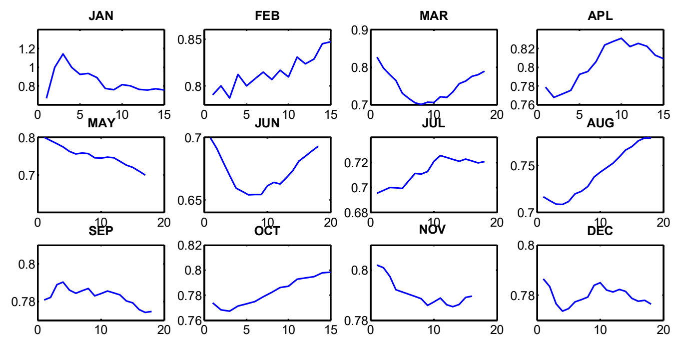

Figure: Tata Motor 2005

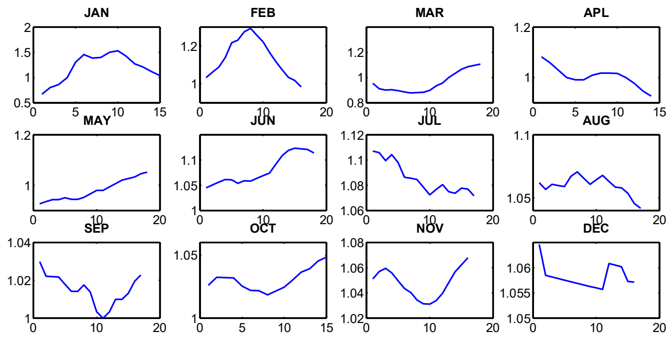

Figure: Tata Motor 2006

Table: Mean Buy/Sell Ratio of Tata Motor from 2004 to 2011

Year	JAN		Mean of FEB to DEC	
	Mean	Std	Mean	Std
2004	—	–	0.7592	0.0063
2005	0.8781	0.2377	0.6224	0.0111
2006	0.7182	0.2840	0.7653	0.0082
2007	0.8854	0.1508	0.7044	0.0071
2008	0.5535	0.1496	0.6212	0.0103
2009	0.9088	0.1182	0.6904	0.0111
2010	0.7403	0.1511	0.7701	0.0077
2011	1.1964	0.2722	1.0440	0.0146

t-Statistics

Tests of difference between the January and Mean of February and December month using t-statistic with n = 18.

$$t = \frac{\bar{X}_d}{S_d/\sqrt{n}}$$

where

$$S_d = \sqrt{\frac{(n-1)*S_1^2 + (n-1)*S_2^2}{2n-2}}$$

H_0 : The difference between January month returns of y^{th} year and Mean of February and December returns of $(y-1)^{th}$ year of Tata Motor Ltd =0

Table: t-Statistic from 2004 to 2011 for Tata Motor

Value of Y	Value of t-statistic
2004	2.778
2005	-**12.831**
2006	4.669
2007	-**58.612**
2008	-5.642
2009	-**50.823**
2010	54.759