PRINCIPAL COMPONENT ANALYSIS TO SELECT INTRADAY STOCKS

S. Calwin Parthibaraj[1]
[1]* Department of Mechanical Engineering, Dr.Sivanthi Aditanar College of Engineering, Tiruchendur, Tamil Nadu, India
e-mail: pscalwin@gmail.com
*Corresponding Author: pscalwin@gmiail.com

Abstract
Intraday trading is a game of gambling in which individual traders invested in a single stock competes with institution traders in volatile and unstable market. In this scenario, the selection of stock is imperative to earn intraday profit is key to win in trading market. Hence in this paper, Principal Component Analysis (PCA) is used to identify the major performance features of companies to select the winning stocks. The results are obtained and interpreted to choose the right stock for the right time.
Keywords: Stock selection, Principal Component Analysis (PCA)

I. INTRODUCTION

In intraday trading, selection of stocks and framing a strategy is a prerequisite to win in trading. In order to select stock the basic fundamental financial structure of the company must be understand thoroughly. As stock is an investment it should be sold at right time without any expectations on its brand and price trend. Hence the selection of stock is made based on the analysis using technical indicators and charts. As many number of charting tools, technical indicators and financial ratios are available, it is cumbersome to assess the stock performance. Basic measures such as liquidity to view the trade volume as a proportion of market capitalisation, opening price gap, post market price gap, current price gap, profit to earning ratio, earnings per share, market capitalization, number of shares traded are available to identify the winning stock. But the correlation between the factors could reduce these selection factors and help to make better decision in stock selection.

Even though, an average score of selection factors would provide a single scale to compare, but with unequal weights the factors could be spread out further on the scale for comparison. But, with the consideration of large number of selection factors, the dispersion matrix may be too large to study and interpret properly. Also there would be too many pairwise correlations between the variables to consider. Hence, it is necessary to reduce the number of variables to a few, interpretable linear combinations of data to interpret it in a more meaningful form.

Principal Component Analysis (PCA) reduces the dimension of these selection factors by using each linear combination of factors into a principal component. The first principal component is the linear combination with maximal variance. It provides a dimension along which the observations are maximally separated or spread out. The second principal component is the linear combination with maximal variance in a direction orthogonal to the first principal component, and so on. Hence, in this paper PCA is considered to select the stock based on the variables that mostly contribute to earn intraday profit. In section 2, the review of literatures is presented, section 3 presents the PCA methodology with assumption and illustration, Section 4 presents the results with discussion, and section 5 concludes with future research.

II. LITERATURE REVIEW

The state of the art in PCA and in stock selection methodology is presented through the review of articles as follows,

Lendasse et al.(2001) reduced the number of technical indicators used to predict and project the market nonlinearly based on PCA.

Cao et al. (2003) compared PCA, Kernels PCA and Independent Component Analysis (ICA) applied to extract features for support vector mechanism in stock forecaster. The comparison resulted that KPCA outperformed other methods in forecasting.

Haron & Maiyastri (2004) classified data into several distinct groups identified the groups using PCA and found better than other ranking methods for modelling stock market returns.

Ince and Trafalis (2007) compared the data mining methods to PCA with support vector regression and multilayer perceptron network to identify the most influential factors for stock price forecast.

Yu et al.(2014) designed a support vector learning machine with PCA of financial ratios to select long term stocks and obtained portfolio significant to share index of shanghai stock exchange.

Sadatrasool et al. (2016) applied PCA with multi criteria decision making process to select manager among competing personnel by extracting effective criteria from all criteria in staff selection.

Pasini (2017) adopted PCA for three subgroups of stocks to optimize stock portfolio for best investment returns.

From the review, it is concluded that the study using PCA were found in designing long term stock portfolio, and for finding factors for forecasting the performance of stocks. However the selection of intraday stock differs from the stocks for long term holding. As intraday stock requires less correlation with market index, high volatility, liquidity, and volume, the basis of stock selection differs. This research gap is identified and fulfilled in this paper by selecting intraday stocks using PCA, and its methodology is presented in next section.

III. PCA METHODOLOGY WITH NUMERICAL ILLUSTRATION

The steps involved in selecting the intraday stock using PCA are presented with an illustration in this section. The assumption that are considered before doing PCA analysis are

- the number of independent variables should be large relative to the number of observations
- the independent variables must be highly correlated, the estimates of regression coefficients may be unstable. In such cases, the independent variables can be reduced to a smaller number of principal components. The input data for the risky penny stock chosen for intraday trading is obtained from Google Finance for PCA. The input data considers six variable measurements on each stock selected in service sector. The details required for PCA are presented in Table 1.

Table 1. Input data for PCA analysis

Company / Index	**Six Measurements**						
	Trade time	**Current gap**	**Open gap**	**Last Traded Price**	**Traded Volume (lacs)**	**P/E**	**EPS**
NSE:DEEPAKFERT	11:28:28 AM	23.05	2.25	189.25	2.96E+06	12.22	15.35
	2:33:55 PM	19.55	2.25	185.75	3.96E+06	12.09	15.35
	2:49:58 PM	18.6	2.25	184.8	4.03E+06	11.99	15.35
	3:03:57 PM	15.55	2.25	181.75	4.25E+06	11.65	15.35

	3:07:58 PM	14.75	2.25	180.95	4.33E+06	11.68	15.35
	3:19:59 PM	11.6	2.25	177.8	4.58E+06	11.56	15.35
	3:27:48 PM	12.45	2.25	178.65	4.67E+06	11.64	15.35
	3:30:00 PM	12.15	2.25	178.35	4.70E+06	11.62	15.35
NSE:MANPASAND	11:28:25 AM	3.3	0.85	98.65	4.68E+05	11.34	8.7
	2:33:58 PM	3.5	0.85	98.85	5.72E+05	11.36	8.7
	2:49:26 PM	3.15	0.85	98.5	5.82E+05	11.24	8.7
	3:03:48 PM	2.2	0.85	97.55	5.96E+05	11.18	8.7
	3:07:53 PM	2.15	0.85	97.5	6.02E+05	11.18	8.7
	3:19:43 PM	2.6	0.85	97.95	6.15E+05	11.21	8.7
	3:27:45 PM	3.4	0.85	98.75	6.64E+05	11.34	8.7
	3:30:00 PM	3.15	0.85	98.5	6.68E+05	11.32	8.7
NSE:WELCORP	11:28:18 AM	2.85	0.5	159.85	2.61E+05	25.35	6.3
	2:33:49 PM	2.4	0.5	159.4	4.72E+05	25.3	6.3
	2:49:58 PM	2.5	0.5	159.5	4.92E+05	25.29	6.3
	3:03:49 PM	2.7	0.5	159.7	5.08E+05	25.4	6.3
	3:07:52 PM	2.5	0.5	159.5	5.20E+05	25.36	6.3
	3:20:02 PM	2.45	0.5	159.45	5.65E+05	25.29	6.3
	3:27:37 PM	2.4	0.5	159.4	5.77E+05	25.31	6.3
	3:30:00 PM	1.7	0.5	158.7	5.82E+05	25.19	6.3

Step 1 : Check PCA assumptions and condition the data

The number of measurements(6) are large relative to the number of observations(8), and the variables are found independently significant and linearly correlated using correlation matrix. As the data is found on different units and scale, logarithmic transformation of input data is made for comparison and presented in Table 2.

Table 2. Conditioning of raw data by data transformation

Company / Index	Trade time	Current gap	Open gap	Last Traded Price	Traded Volume (lacs)	P/E	EPS
NSE:DEEPAKFERT	11:28:28 AM	1.3627	2.25	2.2770	6.4711	1.0871	1.1861
	2:33:55 PM	1.2911	2.25	2.2689	6.5975	1.0824	1.1861
	2:49:58 PM	1.2695	2.25	2.2667	6.6050	1.0788	1.1861
	3:03:57 PM	1.1917	2.25	2.2595	6.6285	1.0663	1.1861
	3:07:58 PM	1.1688	2.25	2.2576	6.6362	1.0674	1.1861
	3:19:59 PM	1.0645	2.25	2.2499	6.6612	1.0630	1.1861

	3:27:48 PM	1.0952	2.25	2.2520	6.6695	1.0660	1.1861
	3:30:00 PM	1.0846	2.25	2.2513	6.6723	1.0652	1.1861
NSE:MANPASAND	11:28:25 AM	0.5185	0.85	1.9941	5.6704	1.0546	0.9395
	2:33:58 PM	0.5441	0.85	1.9950	5.7571	1.0554	0.9395
	2:49:26 PM	0.4983	0.85	1.9934	5.7651	1.0508	0.9395
	3:03:48 PM	0.3424	0.85	1.9892	5.7753	1.0484	0.9395
	3:07:53 PM	0.3324	0.85	1.9890	5.7792	1.0484	0.9395
	3:19:43 PM	0.4150	0.85	1.9910	5.7888	1.0496	0.9395
	3:27:45 PM	0.5315	0.85	1.9945	5.8221	1.0546	0.9395
	3:30:00 PM	0.4983	0.85	1.9934	5.8250	1.0538	0.9395
NSE:WELCORP	11:28:18 AM	0.4548	0.5	2.2037	5.4165	1.4040	0.7993
	2:33:49 PM	0.3802	0.5	2.2025	5.6738	1.4031	0.7993
	2:49:58 PM	0.3979	0.5	2.2028	5.6923	1.4029	0.7993
	3:03:49 PM	0.4314	0.5	2.2033	5.7055	1.4048	0.7993
	3:07:52 PM	0.3979	0.5	2.2028	5.7162	1.4041	0.7993
	3:20:02 PM	0.3892	0.5	2.2026	5.7523	1.4029	0.7993
	3:27:37 PM	0.3802	0.5	2.2025	5.7615	1.4033	0.7993
	3:30:00 PM	0.2304	0.5	2.2006	5.7648	1.4012	6.3
Average		0.6779	1.2000	2.1518	6.0253	1.1758	1.2042

Step 2: Determine variance covariance matrix and total variance

As PCA deals with the measurements that are independent without any groupings statistical test is not needed. Hence the variance-covariance matrix is determined using Excel matrix function(mmult) and the presented in tabular form in Table 3.

Table 3. Determination of variance-covariance matrix

0.146	0.286	0.027	0.156	-0.032	-0.049
0.286	**0.597**	0.047	0.330	-0.078	-0.043
0.027	0.047	**0.014**	0.030	0.007	0.018
0.156	0.330	0.030	**0.190**	-0.038	0.005
-0.032	-0.078	0.007	-0.038	**0.027**	0.034
-0.049	-0.043	0.018	0.005	0.034	**1.203**

The total variance is the sum of the variance along the diagonal in variance covariance matrix and it is given by

Total variance = 0.146 + 0.597 + 0.014 + 0.190 + 0.027 + 1.203 = 2.1768

Step 3 : Determination of Eigenvalues and Eigenvectors to each eigenvalue

As the variance-covariance matrix involves six rows and columns, QR method is employed to determine the eigenvalues and eigenvectors. The six eigenvalues and its corresponding eigenvectors are obtained and presented in Table 4.

Table 4. Eigenvalues and its corresponding eigenvectors

	1	2	3	4	5	6
	Eigenvalue, E1= 0	Eigenvalue, E2 = 0.002	Eigenvalue, E3 = 0.009	Eigenvalue, E4 = 0.029	Eigenvalue, E5 = 0.922	Eigenvalue, E6 = 1.215
Eigen vector f	96.398	65.635	19.445	-9.851	2.025	-0.096
	188.162	-71.459	1.02	5.548	4.264	-0.151
	-1518.658	-23.473	-1.191	-22.077	0.376	0.006
	61.112	683717	-18.936	-4.67	2.402	-0.06
	1136.523	-28.798	-2.392	-29.333	-0.505	0.043
	1	1	1	1	1	1

Step 4: Determination of number of principal components

The number of components that contribute to maximum variance of more than 90% is determined by arranging the eigenvalues in descending order. The cumulative eigenvalues are calculated, and its proportion with respect to total variance is obtained and presented in Table 5.

Table 5. Eigenvalue and its proportion with total variance

Eigenvalue	Proportion of variance	Cumulative proportion
1.215	0.558	0.558
0.922	0.424	0.982
0.029	0.013	0.995
0.009	0.004	0.999
0.002	0.001	1.000
0.000	0.000	1.000

As the eigenvalues, E6 = 1.215 and E5 = 0.922 contributes 98.2 percentage of total variance, the two principal components are sufficient to explain the effect on selection factors on each stock. The first principal component corresponds to eigenvalue (E6 = 1.215) with 55.8 percent variance and the second principal component with eigenvalue (E5 = 0.922) contributes 42.4 percent of total variance with cumulative percent of 98.2 percent.

Step 5: Computation of principal component scores and values to each individual observation.

The reduction in dimensions of original variables by PCA, and the magnitude or weightage for each selection factors is calculated using eigenvectors for each eigenvalue. The coefficient for the first principal component that is the eigenvector for eigenvalue (E6) is given by

Table 6. Coefficients for principal components

Variables	First Principal Component, Z1	Second Principal Component, Z2
Current gap	-0.096	2.025
Open gap	-0.151	4.264
Last Traded Price	0.006	0.376
Traded Volume (lacs)	-0.06	2.402
P/E	**0.043**	-0.505
EPS	**1**	1

From Table 6 the first principal component corresponding to eigenvalue of 1.215 is given by

Z1 = -0.096 X Current gap -0.151 X Open gap + 0.006 X Last Traded Price - 0.06 X Current Traded Volume + 0.043 + EPS and the second principal component corresponding to eigenvalue of 0.922 is given by

Z2 = 2.025 X Current gap +4.264 X Open gap + 0.376 X Last Traded Price + 2.402 X Current Traded Volume -0.505 + EPS

By substituting the original values to the principal component, the principal component scores are computed and presented in Table 7.

Table 7. Principal component scores for the principal components

Company / Index	Last trade time	Z1	Z2
NSE:DEEPAKFERT	11:28:28 AM	**0.3877**	**29.3903**
	2:33:55 PM	**0.3867**	**29.5483**
	2:49:58 PM	**0.3882**	**29.5235**
	3:03:57 PM	**0.3937**	**29.4260**
	3:07:58 PM	**0.3954**	**29.3969**
	3:19:59 PM	**0.4037**	**29.2451**
	3:27:48 PM	**0.4004**	**29.3265**
	3:30:00 PM	**0.4012**	**29.3118**
NSE:MANPASAND	11:28:25 AM	**0.4785**	**19.4515**
	2:33:58 PM	**0.4709**	**19.7113**
	2:49:26 PM	**0.4746**	**19.6396**
	3:03:48 PM	**0.4888**	**19.3482**
	3:07:53 PM	**0.4895**	**19.3373**
	3:19:43 PM	**0.4811**	**19.5274**
	3:27:45 PM	**0.4681**	**19.8422**
	3:30:00 PM	**0.4711**	**19.7821**
NSE:WELCORP	11:28:18 AM	**0.4288**	**16.9823**
	2:33:49 PM	**0.4205**	**17.4493**
	2:49:58 PM	**0.4176**	**17.5299**
	3:03:49 PM	**0.4137**	**17.6284**
	3:07:52 PM	**0.4163**	**17.5867**
	3:20:02 PM	**0.4149**	**17.6561**
	3:27:37 PM	**0.4152**	**17.6599**
	3:30:00 PM	**0.4293**	**17.3650**

Step 6 : Determination of correlation of Principal component with the original variables

The principal component variable is added to the six original variables and the correlation matrix is obtained and presented in Table 8.

Table 8. Correlation between variables including principal component

Variables	**Current gap**	**Open gap**	**Traded Price**	**Traded Volume (lacs)**	**P/E**	**EPS**	**Z1**	**Z2**
Current gap	1.000	0.970	0.593	0.939	-0.511	-0.116	**-0.674**	**0.975**
Open gap	0.970	1.000	0.513	0.979	-0.615	-0.051	**-0.578**	**0.999**
Last Traded Price or close	0.593	0.513	1.000	0.587	0.361	0.143	**-0.983**	**0.534**
Current Traded Volume (lacs)	0.939	0.979	0.587	1.000	-0.523	0.011	**-0.651**	**0.984**
P/E	-0.511	-0.615	0.361	-0.523	1.000	0.188	**-0.274**	**-0.595**
EPS	-0.116	-0.051	0.143	0.011	0.188	1.000	**-0.072**	**-0.051**
Z1	-0.674	-0.578	-0.983	-0.651	-0.274	-0.072	**1.000**	**-0.603**
Z2	0.975	0.999	0.534	0.984	-0.595	-0.051	**-0.603**	**1.000**

The coefficient of principal components and the correlation between principal component and original variables obtained in this section are discussed and interpreted in Section 4.

IV. Results and inferences

The first two principal components are

Z1 = -0.096 X Current gap -0.151 X Open gap + 0.006 X Last Traded Price - 0.06 X Current Traded Volume + 0.043 + EPS

Z2 = 2.025 X Current gap +4.264 X Open gap + 0.376 X Last Traded Price + 2.402 X Current Traded Volume -0.505 + EPS

It is noted that the coefficients of current price gap, open price gap, traded volume and Earnings Per Share are higher and have notable influence on principal components. It also corresponds to larger variances along the diagonal of variance-covariance matrix. Further the interpretation of correlation matrix for principal components presented in Table 8 need to be analysed. The interpretation will be based on the conditions that is the larger and farther values from zero with values greater than or less than 0.7 are deemed to be important for consideration in stock selection. These factors are bolded in Table 9 and it represents the influence of principal components with the original variable. It shows that the first principal component decreases with only one variable that is closing price, and the second principal component increase with three variables that is current price gap, traded volume and open price gap.

Table 9. Importance of first and second principal component

Variables	Z1	Z2
Current gap	-0.674	**0.975**
Open gap	-0.578	**0.999**
Last Traded Price or close	**-0.983**	0.534
Current Traded Volume (lacs)	-0.651	**0.984**
P/E	-0.274	-0.595
EPS	-0.072	-0.051

V. Conclusion and future directions

The intraday stock market is affected by various factors that are explanatory and important to be considered to conduct market research. Hence in this paper, the factors that mostly influence the traders in selecting a stock are considered to perform descriptive PCA of these data. The results presents that the first principal component is strongly correlated with one of the original variables. The first principal component decreases with increasing scores of last traded price. This component can be viewed as a measure of the variation in market due to last trading price. And the second principal component increases with three of the values such as opening price gap, trading volume and current price gap. This second component can be viewed as a measure of how traders vary in strategic approach to win in the trading market. Thus the interpretation points out that the selection of stock should be made based on the significant factors considered in the order of last trade price, opening price gap, trading volume and current price gap. PCA guide to make a decision considering data of important. But the problem with PCA is that the interpretation could not be done if the same factors are considered important for two or more principal components that opens the scope for future research.

References

[1] Cao, L., Chua, K., Chong, W., Lee, H., & Gu, Q. (2003, 09). A comparison of PCA, KPCA and ICA for dimensionality reduction in support vector machine. *Neurocomputing*, *55*(1-2), 321-336.

[2] Feeney, G. J., & Hester, D. D. (1964). *Stock market indices: a principal components analysis* (No. 175). Cowles Foundation for Research in Economics, Yale University.

[3] Haron, K. (2004). Principal component analysis in modelling stock market returns. *Matematika*, *20*(1), 31-41.

[4] Ince, H., & Trafalis, T. B. (2007, 03). Kernel principal component analysis and support vector machines for stock price prediction. *IIE Transactions*, *39*(6), 629-

637.

[5] Lendasse, A., Lee, J., Bodt, D., Wertz, V., & Verleysen, M. (2001). Dimension reduction of technical indicators for the prediction of financial time series - Application to the BEL20 Market Index. *European Journal of Economic and Social Systems, 15*(2), 31-48.

[6] Pasini, G. (2017, 06). Principal Component Analysis For Stock Portfolio Management. *International Journal of Pure and Apllied Mathematics*, *115*(1).

[7] Sadatrasool, M., Bozorgi-Amiri, A., & Yousefi-Babadi, A. (2016). Project manager selection based on project manager competency model: PCA–MCDM Approach. *Journal of Project Management,* (10), 7-20.

[8] Yu, H., Chen, R., & Zhang, G. (2014). A SVM Stock Selection Model within PCA. *Procedia Computer Science*, *31*, 406-412.