ORIGINAL PAPER

DOI: 10.26794/2587-5671-2023-27-5-115-127
JEL G11, G12, G15

Data Mining in Indian Equity Markets: Building Low Risk, Market Beating Portfolios

S.R. Mitragotri, N. Patel
Nirma University, Ahmedabad, India

ABSTRACT
Over the last five decades, business academics have identified over 300 determinants that potentially influence stock returns. However, we still do not know whether all return determinants are equally important, or whether there is a smaller set of determinants that has a disproportionately larger influence on stock returns. Can mining historical data help us find this smaller set of return determinants that has a disproportionately higher influence on stock returns? Using historical data from the Indian market, we build a large database of investments with more than 74,000 investments spread over a period of 132 months. From this database, using "association rule mining" method, we are able to mine a strong set of "association rules" that point to a smaller set of "return determinants" that are seen more frequently in investments that beat index returns. From a pool of thirty-seven return determinants, using "association rule mining", we were able to find out a small set of key return determinants that are seen most frequently in investments that beat index returns in India. Portfolios created from these "association rules" have a portfolio risk lower than the market risk and provide index-beating returns. "Out-of-sample" portfolios created using these association rules have portfolio "Beta" less than one and provide returns that beat the market returns by a significant margin for all holding periods in the Indian market. Through this paper, we demonstrate how portfolio managers can mine "association rules" and build portfolios without any limits on the number of factors that can be included in the screening process.
Keywords**:** stock returns; mining association rules; return-determinants; portfolio-risk

For citation: **Mitragotri S.R., Patel N. Data mining in Indian equity markets: Building low risk, market beating portfolios. *Finance: Theory and Practice*. 2023;27(5):115-127. DOI: 10.26794/2587-5671-2023-27-5-115-127**

INTRODUCTION

Since 1970, business academics have identified more than 330 firm level return determinants [1]. However, we still have unanswered questions such as: are all these return determinants equally important in predicting stock returns? In this large pile of return determinants, is there a smaller set of return determinants with a stronger ability to predict stock returns? If there indeed is such a smaller set of return determinants — how do we uncover them? Can mining historical data help us answer these questions?

This study places historical data on stock returns and 37 highly prevalent return determinants in a single frame, and, with the help of "association rule mining" successfully, identifies return determinants that are seen more frequently in index-beating investments. Portfolios built from the mined association rules have a lower risk than the market, and yield returns that are significantly above market returns and perform equally well in an out-of-sample data set.

The remainder of this paper is organized as follows: in Section 2, we look at past research on factors influencing stock returns and try to understand the reasons for some of the contradicting inferences about factors influencing stock returns. We also examine recent methodologies used in empirical asset pricing research including the application of analytics and machine learning techniques. Section 3 briefly explains the association mining technique and analyzes the association rules mined between stock returns and return determinants. Section 4 tests the mined association rules and Section 5 concludes the paper.

LITERATURE REVIEW

In this section, we highlight how factors proven to be strong return determinants in one study are challenged, and proven to be insignificant in a subsequent one. The objective is to uncover the possible reasons for such contradicting inferences about the factors that influence stock returns.

Numerous studies have been carried out to identify factors that influence stock returns. One of the highly studied return determinants is the P/E ratio [2–4]. The predominant observation from these studies is that portfolios with low P/E stocks have lower systematic risk and earn significantly higher returns compared to portfolios with higher P/E stocks. Another factor that has been closely studied for its impact on stock returns has been the "debt/equity" ratio (D/E) [5–7]. The D/E ratio of a company is a useful proxy for risk and a higher D/E ratio indicates a higher degree of risk for equity holders, which is seen in higher expected stock returns. However, another study [8] finds that the impact of these two factors, P/E and D/E, is subsumed by two other factors: size and BV/P (book value/price). A study by W.C. Barbee et al. [6] challenged the role of BV/P and size in predicting stock returns. Instead, they find that sales-price ratio and D/E ratio explain stock returns better than BV/P or size. This study reports that the sales-price ratio also captures the role of the D/E ratio in explaining stock returns, thus making the sales-price ratio a more reliable return determinant.

This cycle of published return determinants being challenged and new return determinants being proposed continues even today. For example, R. Alquist et al. [9] challenged the impact of size on stock returns. They report that while the "size effect" is seen in the market, returns to size are neither persistent nor stable; hence it is not a key factor for constructing portfolios. More recently, R. Ball et al. [10] argue that P/BV is a good predictor of stock returns because the retained earnings part of book value aggregates past earnings, which is a strong indicator of a firm's earnings history. They further report that retained earnings/price is a good predictor of returns and that contributed capital has no ability to predict stock returns.

Asset pricing research is now at a stage, where approximately 18 new factors are discovered annually [11] creating what J.H. Cochrane [12] calls a "zoo" of factors. For example, new factors being studied for their impact on asset returns relate to the environmental impact, social impact and governance (ESG) of the organization [13–15]. C.R. Harvey et al. [11] have identified 316 factors from top journals and believe that this probably underrepresents the factor population.

One reason for this tussle between different return determinants in different studies can be attributed to the choice of linear regression as a method used in these studies. When linear regression is applied to understand the relationship between stock returns and return determinants, it is very difficult to include more than four return determinants in a single study [16]. This leads to a situation where a researcher selects four factors and identifies a couple of strong factors as "key return determinants". The next research considers these "key return determinants" along with a few other factors and proves that the first two return determinants do not influence as much as the new set of factors in a different period of study. We see this happening repeatedly in a large part of asset pricing research over the last five decades. Given the fact that we have more than 300 documented return determinants, we need to use a method that will allow us to include as many potential return determinants as possible in the same study and understand the strength of each one's influence on stock returns.

Towards that objective, we see a lot of interesting studies that use different methodologies to understand the influence of factors on stock returns. E.H. Sorensen [17] traces the evolution of quantitative methods in investing and portfolio management including recent machine learning techniques. X. Wu et al. [18] have used both multivariate regression and a novel machine learning models to examine the effect of expert analysts' recommendations on stock prices. Y. Li and Y. Pan [19] have developed an ensemble of deep learning model to predict future stock prices. K.C. Rasekhschaffe and R.C. Jones [20] provide a very good introduction to machine learning algorithms. Multiple literature reviews focus on prior work that applied machine learning to empirical asset pricing and portfolio management [21, 22].

In this paper, we use a data-mining technique called "mining association rules" to explore the relationship between stock returns and return determinants. In the next section, we examine the framework for mining association rules between stock returns and a large pool of return determinants.

MINING ASSOCIATION RULES BETWEEN STOCK RETURNS AND RETURN DETERMINANTS

Mining for association rules between different variables in a large database is widely adopted in industries that generate multidimensional data. We briefly look at what "association rule mining" is and how it can be used to mine association insights between stock returns and return determinants.

Mining "association rules" involves identifying item clusters in a database. For example, in the retail industry, this technique is used to discover groups of products that tend to be purchased together. In our study, the item cluster we are looking for is a "set of return determinants" regularly observed in index-beating investments.

Information about the associations mined is expressed in form of "if-then" statements that are probabilistic in nature. For example, an association rule mined from the transaction database of a retail store could be: "If a buyer has purchased milk and butter, then there is 80% probability that she/he will also buy bread". This is inferred from the actual number of transactions recorded in the database. This means that of the 100 customers who had purchased milk and butter, 80 of them had also purchased bread. This is how the "if-then" association rules are formed based on historical transactional data. A. Rai [23] provides a very good overview of mining association rules.

To build a database to mine association rules, we would need data on different return determinants at the time of investment and data about stock returns and index returns for different holding periods after the investment is made. Imagine an investor who invests in a large set of stocks on the 1st May 2002 and continues to invest every month on the same date in the same set of stocks for the next ten years. Every month, at the time of investment, for each of his investments, he has data of company-reported information about different return determinants like sales, earnings, P/BV, etc. This information about different return determinants at the time of investing is the first part of the database. For investments made at different points in time, we obtain data about actual returns relative to the index returns for different holding periods. Such a database will enable us to mine the association rules between return determinants at the time of investing and stock returns for different holding periods.

Building the Database to Mine Association Rules

The first task in building the database for mining association rules was to identify the return determinants. Researchers have identified over 300 factors that impact stock returns. Although association rule mining does not limit the number of return determinants that can be included in the study, we considered thirty-seven return determinants that are considered important in fundamental analysis i.e., accounting data, which are considered strong predictors of stock prices. We did not consider technical indicators because our primary intent was to mine associations between factors and returns for longer holding periods.

Data

For this study, we have taken data from companies listed on the National Stock Exchange of India (NSE) and Bombay Stock Exchange (BSE). The data source was "Refinitiv Datastream". The thirty-seven return determinants considered for this study are listed in Appendix (*Table 1*).

We considered monthly investments from January 2002 to December 2012, with investments made on the first of every month. This created a pool of 74,869 investments spread over 132 months. We mined association rules between return determinants and stock returns for holding periods of one, three and five years.

The market index considered for computing market returns is NSE Nifty 50.

Based on the above information, a comprehensive database was created to mine associations rules.

Associations Mined

We used libraries available in R-programming language to mine the association rules. From the large set of association rules mined, we considered ten strong association rules for analysis and they are listed in Appendix (*Table 2*).

Interpreting the Association Rules

Consider association rule No. 2 for 3 year holding period shown in *Table 1* below.

'LHS' (Left Hand Side) in the above table is the 'If' part of the 'If-Then' statement of the association rule.

'RHS' (Right Hand Side) in the above table is the 'Then' part of the 'If-Then' statement of the association rule.

The confidence column is the confidence of the association rule.

Let us interpret this association rule — it says:

"*If, for a stock:*

- The "price divided by the sales-per-share is less than 1", AND "Debt by Working capital is less than 2", AND 'Return on Invested Capital has consistently been above 12% in the last 5 years' AND it has very high "Cash Flow by Assets" (in top 33 percentile)

Then

There is an 75.2% chance that the "<u>3 years returns</u>" from investing in that stock will be greater than the "<u>3 years index returns</u>" for the same holding period.

The above "If-then" statement is based on the first three columns of the association rule.

We can also see that this association rule is based on performance of 848 investments in the investment database created for this study. The lift value greater than

Table 1

Example of Association Mined

LHS (Antecedent)	Confidence	RHS (Consequent)	Support	Count	Lift
PriceBySalesPerShareLT1, TTMDbyWCLT2, T5RoICGT12, T33CFbyAssets	75.2%	n3YrRtn_GTNSE 50	1.1%	848	1.81

Source: Compiled by the authors.

one (1.81) confirms that this is a strong association rule and not a chance occurrence.

Based on the rules considered for analysis in this study, we can make the following observations:

- Index-beating returns can be achieved by picking stocks based on different return determinants in different combinations. For example, in association rule number 1, we see that investment in stocks with certain return determinants (mentioned in the "If" part of the rule) have a high probability of yielding market-beating returns. Association rule No. 2 has a completely different combination of return determinants with a high probability of index-beating returns. This implies that there are multiple ways to achieve market-beating returns and raises questions about the validity of the quest for a single asset pricing model based on a fixed set of factors.
- There are no strong association rules for a 1-year holding period (confidence lower than 70%).
- In almost all cases, we see that the confidence of the association is comparatively higher for a 5-year holding period. This implies that for smaller holding periods, it is difficult to find strong associations between return determinants and market-beating returns. However, for longer holding periods (three to five years), we find strong association rules between return determinants and market-beating returns.

Key Return Determinants

From all the strong associations mined and tabulated in *Appendix* (*Table 2*), we listed return determinants that appear most frequently in these association rules. We consider these factors as key return determinants — factors key for predicting index-beating stock returns. The key return determinants are listed below:

1) T33_T5AvgSalesByT5AvgAssets: "Average of past five-years of sales by average of past five-years of assets" is in top 33 percentile among all the investment opportunities considered.

2) T5RoICGT12: For every year in the last 5 years, the Return on Invested Capital was greater than 12%.

3) T33CFbyAssets: Cash flow by assets is in the top 33 percentile among all investment opportunities considered. (Cash flow = earnings + depreciation).

4) T5SalesGrowthGT1.05: Last five-years year-on-year sales growth is more than 5%.

5) T5BVGrowthGT1: Each year in the last 5 years, year-on-year growth in Book Value is greater than 1.

6) TTMDbyWCLT2: Debt by Working Capital is less than 2.

7) TTMDERatioLE 1: D/E Ratio is less than or equal to 1.

8) T33SalesByRcvbl: Sales by Accounts Receivables is in the top 33 percentile among all the investment opportunities considered.

The above factors and their strong association with market-beating returns convey that if the stock you are investing in has a certain combination of the above factors, then there is a very good probability that such an investment will yield market-beating returns for holding periods ranging between one and five years. The right combination of the above factors for market-beating returns can be seen in the LHS of the association rules listed in *Appendix* (*Table 2*).

The above key return determinants give some interesting insights:

- The first 3 metrics emphasize the importance of capital efficiency. Firms that deploy capital more efficiently than their peers show better results, which is reflected in the index-beating returns from investment in those stocks.
- The next 2 metrics emphasize growth and this is important for a growing economy like India.
- The next 2 metrics underscore the need for limiting the "debt" of the company to reasonable levels.

VALIDATING THE ASSOCIATIONS MINED

We have tested the associations mined in two different ways as described below:

1. Using an out-of-sample data set, we compute the risk of "association rule portfolios" and compare the risk-return of these portfolios with the market returns.

2. The second validation of the association rules was to check the performance of these association rules when linear regression methods were applied to them.

Risk Adjusted Returns

First, we compute the portfolio risk of association rule portfolios. For that, we created "association rule portfolio" for all ten association rules analyzed in this study. The "association rule portfolio" comprises stocks that meet the LHS criteria ("If" part) of the association rule.

For constructing association rule portfolios and computing the portfolio betas, we collected the below data from "Refinitiv Datastream":

Data on monthly stock price and return determinants from Jan-2013 to Dec-2014 for 832 companies listed on NSE-India and BSE-India. The data set required to compute the 5-year returns of an investment made in Dec-14 extends up to Dec-19. So, the period covered in this study extends up to Dec-19. We did not consider the period beyond 2019 to ensure that our findings were not influenced by the uncertain economic period of the 2020 global pandemic.

NSE Nifty-50 (Index) data for the above period to compute the market returns.

To compute risk-free returns, we consider the 91-day government treasury bill yield as the risk-free rate.

To compute portfolio beta, we use the following equation provided by CAPM:

$$R_p - R_f = \beta(R_m - R_f).$$

We ran OLS regression with dependent variable as "$R_p - R_f$" and independent variable as "$R_m - R_f$" to estimate portfolio β. The value of the portfolio beta for the portfolios of each association rule is shown below in *Table 2*.

As seen in *Table 2*, portfolio β for every association rule considered in this study is less than 1, which means that the association rule portfolios have a lower risk level than the market risk.

For all ten association rules analyzed here, we constructed price-weighted portfolios every month between Jan-13 and Dec-14. This gave 24 portfolios for each association rule. We compared the portfolio returns of these 24 portfolios for each association rule with the index returns for the corresponding holding period.

The performance of these portfolios for different association rules created at different points in time is tabulated in *Appendix* (*Table 3*). Below are some important observations:

- For 3-year holding period, for all 10 association rules and all 24 monthly portfolios, the portfolio returns are greater than the index returns — at lower risk than the market.
- For 5-year holding period, in 9 of the 10 association rules, the portfolio returns for all 24 monthly portfolios were greater than the index returns. For one association rule (association rule No. 2), returns of one of the 24 portfolios is lower than the index returns — which is a very small percentage of failure of the association rule.
- The association rules considered here are not strong for 1-year holding period. However, we analyzed portfolio returns for 1-year holding period for all ten association rule portfolios. We find that in this case as well, in 67% or more cases, portfolio returns are higher than index returns. Therefore, the association rules performed reasonably well even when they were not very strong.

Association Rules and Regression

To verify performance of association rules in the regression model, we used the LOGIT model because both the antecedent and consequent of the association rules are binary in nature. For the LOGIT regression, we consider the antecedent(s) of the association rule as the independent variables and the consequent part of the association rule as the dependent variable.

We find that the parameters are statistically significant for eight of the ten rules being analyzed. These rules also passed the following model consistency tests:

- Likelihood ratio test
- Wald Test
- Variance Inflation Factor Test

This outcome of the logit regression when applied to the association rules raises an interesting question. Does the real value of a return determinant matter in driving the performance of stock returns relative to the index? Or is the value of the return determinant above or below a certain threshold more important in determining the

Table 2

Portfolio Beta for Different Association Rule Portfolios

Association Rule Portfolio ↓	Portfolio Beta
Rule 1	0.94
Rule 2	0.89
Rule 3	0.888
Rule 4	0.91
Rule 5	0.93
Rule 6	0.89
Rule 7	0.92
Rule 8	0.97
Rule 9	0.81
Rule 10	0.85

Source: Compiled by the author.

stock performance? For e.g.— does the actual value of Price / Sales matter, or is its value above or below the threshold of 1 more important in determining the stock's relative performance?

CONCLUSION

We began this research by looking for answers to a few questions related to asset pricing: In a pile of over 300 potential return determinants, is there a smaller set of return determinants that can be a stronger predictor of stock returns? Is there a way to uncover that set of key return determinants?

This study has largely been able to answer these questions. Takeaways from this study are as follows:

1. Using association mining method, from a pool of 37 return determinants, we were able to extract a smaller set of 8 return determinants that are seen most frequently in investments with market-beating returns. These return determinants are:
   a. Past 5 years Y-O-Y sales growth greater than 5%;
   b. High value of Sales/Account Receivables;
   c. High value of '5 years average of sales/5 years average of assets';
   d. Debt / Working Capital less than 2;
   e. Debt/Equity less than 1;
   f. Past 5 years RoIC greater than 12%;
   g. Year-on-year positive increase in book value;
   h. High value of ratio "Cash Flow / Assets".
2. In an out-of-sample data set, portfolios created from these association rules have portfolio "beta" less than one and provide returns that beat the market returns by a significant margin for all holding periods.
3. Portfolio managers can use the association mining process to identify strong associations between the factors of their choice and index-beating returns.
4. Finally, when we applied the LOGIT model to the association rules, we found that the coefficients were statistically significant for eight out of the ten association rules analyzed.

Data Availability

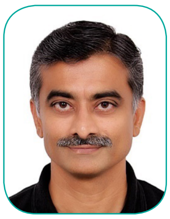
The data supporting the findings of this study are available in the general public repository "Figshare" at DOI: https://doi.org/10.6084/m9.figshare.21399549

REFERENCES

1. Green J., Hand J., Zhang F. The remarkable multidimensionality in the cross-section of expected US stock returns. *SSRN Electronic Journal*. 2014. DOI: 10.2139/ssrn.2262374
2. Basu S. Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis. *The Journal of Finance*. 1977;32(3):663–682. DOI: 10.1111/j.1540–6261.1977.tb01979.x
3. Trevino R., Robertson F. P/E ratios and stock market returns. *Journal of Financial Planning*. 2002;15(2):76–84.
4. Park S. The P/E ratio, the business cycle, and timing the stock market. *The Journal of Portfolio Management*. 2021;47(8):165–183. DOI: 10.3905/jpm.2021.1.270
5. Bhandari L.C. Debt/equity ratio and expected common stock returns: Empirical evidence. *The Journal of Finance*. 1988;43(2):507–528. DOI: 10.1111/j.1540–6261.1988.tb03952.x
6. Barbee W.C., Jr., Mukherji S., Raines G.A. Do sales-price and debt-equity explain stock returns better than book-market and firm size? *Financial Analysts Journal*. 1996;52(2):56–60. DOI: 10.2469/faj.v52.n2.1980
7. Obreja I. Book-to-market equity, financial leverage, and the cross-section of stock returns. *The Review of Financial Studies*. 2013;26(5):1146–1189. DOI: 10.1093/rfs/hhs180
8. Fama E.F., French K.R. The cross-section of expected stock returns. *The Journal of Finance*. 1992;47(2):427–465. DOI: 10.1111/j.1540–6261.1992.tb04398.x

9. Alquist R., Israel R., Moskowitz T. Fact, fiction, and the size effect. *The Journal of Portfolio Management*. 2018;45(1):34–61. DOI: 10.3905/jpm.2018.1.082
10. Ball R., Gerakos J., Linnainmaa J.T., Nikolaev V. Earnings, retained earnings, and book-to-market in the cross section of expected returns. *Journal of Financial Economics*. 2020;135(1):231–254. DOI: 10.1016/j.jfineco.2019.05.013
11. Harvey C.R., Liu Y., Zhu H. ... and the cross-section of expected returns. *The Review of Financial Studies*. 2016;29(1):5–68. DOI: 10.1093/rfs/hhv059
12. Cochrane J.H. Presidential address: Discount rates. *The Journal of Finance*. 2011;66(4):1047–1108. DOI: 10.1111/j.1540–6261.2011.01671.x
13. Efimova O.V., Volkov M.A., Koroleva D.A. The impact of ESG factors on asset returns: Empirical research. *Finance: Theory and Practice*. 2021;25(4):82–97. DOI: 10.26794/2587–5671–2021–25–4–82–97
14. Breedt A., Ciliberti S., Gualdi S., Seager P. Is ESG an equity factor or just an investment guide? *The Journal of Investing*. 2019;28(2):32–42. DOI: 10.3905/joi.2019.28.2.032
15. Giese G., Lee L.-E., Melas D., Nagy Z., Nishikawa L. Foundations of ESG investing: How ESG affects equity valuation, risk, and performance. *The Journal of Portfolio Management*. 2019;45(5):69–83. DOI: 10.3905/jpm.2019.45.5.069
16. Goyal A. Empirical cross-sectional asset pricing: A survey. *Financial Markets and Portfolio Management*. 2012;26(1):3–38. DOI: 10.1007/s11408–011–0177–7
17. Sorensen E.H. The golden age of quant. *The Journal of Portfolio Management*. 2019;46(1):12–24. URL: https://www.panagora.com/wp-content/uploads/JPM-The-Golden-Age-of-Quant.pdf
18. Wu X., Ye Q., Hong H., Li Y. Stock selection model based on machine learning with wisdom of experts and crowds. *IEEE Intelligent Systems.* 2020;35(2):54–64. DOI: 10.1109/MIS.2020.2973626
19. Li Y., Pan Y. A novel ensemble deep learning model for stock prediction based on stock prices and news. *International Journal of Data Science and Analytics*. 2022;13(2):139–149. DOI: 10.1007/s41060–021–00279–9
20. Rasekhschaffe K.C., Jones R.C. Machine learning for stock selection. *Financial Analysts Journal*. 2019;75(3):70–88. DOI: 10.1080/0015198X.2019.1596678
21. Weigand A. Machine learning in empirical asset pricing. *Financial Markets and Portfolio Management*. 2019;33(1):93–104. DOI: 10.1007/s11408–019–00326–3
22. Bartram S.M., Branke J., De Rossi G., Motahari M. Machine learning for active portfolio management. *The Journal of Financial Data Science*. 2021;3(3):9–30. DOI: 10.3905/jfds.2021.1.071
23. Rai A. An overview of association rule mining & its applications. upGrad Education Private Limited. Sep. 28, 2022. URL: https://www.upgrad.com/blog/association-rule-mining-an-overview-and-its-applications/

ABOUT THE AUTHORS



Srinath R. Mitragotri — Doctoral Student, Institute of Management, Nirma University, Ahmedabad, India
http://orcid.org/0000-0001-9493-280X
Corresponding author:
Srinath.mitragotri@gmail.com

Nikunj Patel — Assoc. Prof., Institute of Management, Nirma University, Ahmedabad, India
http://orcid.org/0000-0003-0693-3349
nikunj@nirmauni.ac.in

Conflicts of Interest Statement: The authors have no conflicts of interest to declare.

The article was submitted on 07.01.2023; revised on 07.03.2023 and accepted for publication on 26.03.2023. The authors read and approved the final version of the manuscript.

APPENDIX

Table 1

List of Return Determinants Considered in this Study

No.	Return Determinant	Variable name used in this paper	Brief Explanation of variable-all variables are binary (Y/N)
1	PAT Margins > 8%	T5PATMarginsGT8	Is the "Profit After Tax" margin > 8% in each of the last 5 years?
2	PAT Margins > 10%	T5PATMarginsGT10	Is the "Profit After Tax" margin > 10% in each of the last 5 years?
3	EPS Growth	T5EPSGrowthGT1	In each of the last 5 Years, is EPS in year 'N' > EPS in year 'N – 1'?
4	Sales Growth	T5SalesGrowthGT1	In each of the last 5 Years, is Sales in year N > sales in year ''N – 1'?
5	Sales Growth	T5SalesGrowthGT1.05	In each of the last 5 Years, is Sales of year N divided by Sales of year 'N – 1' > 1.05?
6	P/Sales	PriceBySalesPerShareLT1	Is Price divided by latest Sales per Share < 1?
7	FCFF	T5FCFF_Positive	Is Free Cash flow to the firm > 0 each year in last 5 Years?
8	Book Value Growth	T5BVGrowthGT1	In each of the last 5 Years, Book Value in year N been > Book Value in year N – 1?
9	P/BV	PriceByTTMBVLE 1	Is Price/ Book Value per Share <= 1?
10	Debt/Working-Capital	TTMDbyWCLT2	Is Debt/Working Capital < 2?
11	D/E Ratio	TTMDERatioLE 1	Is Debt/Equity <= 1?
12	P/E Ratio	PERatioLE 10	Is Price/EPS <= 10?
13	PE Ratio/EPS Growth	TTMPEGLE 1	Is Price/EPS ratio divided by EPS growth in Percent <= 1?
14	Return on Invested Capital (ROIC)	T5RoICGT12	In each of the last 5 Years, is RoIC > 12%?
15	Return on Invested Capital	T5RoICGT15	In each of the last 5 Years, is RoIC > 15%?
16	EPS/P + DPS/P + (EPS-DPS)/BVPS	ThumbRuleGE 0.25	Is ThumbRule value >= than 0.25? ThumbRule = (EPS/Price) + (Div. per Share / Price) + ((EPS – Div. per Share) / Book Value per Share)
17	Dividend Yield	T33AvgT3DY	Is Dividend yield in the top 33 percentile amongst all investment opportunities considered?
18	EBIT / EV	T33TTMEBITbyEV	Is EBIT/EV in the top 33 percentile amongst all investment opportunities considered?
19	Gross Profit/Assets	T33TTM_GrProfitByAssets	Is Gross Profit divided by Total Assets in top 33 percentile amongst all the investment opportunities considered?
20	EBIT / Assets	T33TTM_EBITByAssets	Is EBIT by Total Assets in the top 33 percentile amongst all the investment opportunities considered?
21	CF / Price	T33TTM_CFperShareByPrice	Is Cash Flow/Price in the top 33 percentile amongst all the investment opportunities considered? (Cash Flow = Earnings + Depreciation)

Table 1 (continued)

No.	Return Determinant	Variable name used in this paper	Brief Explanation of variable-all variables are binary (Y/N)
22	CF / Assets	T33CFbyAssets	Is Cash Flow/Assets in the top 33 percentile amongst all the investment opportunities considered?
23	FCFF / Assets	T33FCFFbyAssets	Is Free Cash Flow to the Firm / Assets in the top 33 percentile amongst all the investment opportunities considered?
24	Sales / Cash	T33SalesByCash	Is Sales / Cash in the top 33 percentile amongst all the investment opportunities considered?
25	Sales / Ac-Receivables	T33SalesByRcvbl	Is Sales / Accounts Receivables in the top 33 percentile amongst all the investment opportunities considered?
26	Sales / Inventory	T33SalesByInventory	Is Sales / Total Inventory in the top 33 percentile amongst all the investment opportunities considered?
27	Debt / CashFlow	B 33TTM_DebtByCF	Is Debt / Cash Flow in the Bottom 33 percentile amongst all the investment opportunities considered?
28	Working-Capital / Sales	B 33TTM_WCbySales	Is Working Capital/Sales in the Bottom 33 percentile amongst all the investment opportunities considered?
29	% Change in Sales > % Change in Inventory	pcChgInSalesGTpcChgInInvtry	Is percent change in sales over previous year > percent change in inventory over previous year?
30	% Change in Sales > % Change in Receivables	TF_ChgInSalesGTRcvbls	Is percent change in sales over previous year > percent change in receivables over previous year?
31	Year-on-Year Asset Growth	B 33_AssetGrowthYoY	Is year on year asset growth in the bottom 33 percentile amongst all the investment opportunities considered?
32	Return on Assets	T33_T5AvgRoA	Is average of past 5 years of return on assets in the top 33 percentile amongst all the investment opportunities considered?
33	Sales / Assets	T33_ T5AvgSalesByT5AvgAssets	Is average of past 5 years of sales divided by average of past 5 years of assets in the top 33 percentile amongst all the investment opportunities considered?
34	Return on Equity	T33_T5AvgRoE	Is average of past 5 years of RoE in the top 33 percentile amongst all the investment opportunities considered?
35	PAT Margins	T33_T5AvgPATMargins	Is average of past 5 years of PAT margins in the top 33 percentile amongst all the investment opportunities considered?
36	E / P Ratio	T33TTM_EPRatio	Is EPS / Price in the top 33 percentile amongst all the investment opportunities considered?
37	Return on Invested Capital	T33_T5AvgRoIC	Is average of past 5 years of RoIC in the top 33 percentile amongst all the investment opportunities considered?

Source: Compiled by the author.

Table 2

List of Strong Association Rules Mined in India & Analyzed

Antecedent ('If' part of the association rule – LHS)	Consequent ("Then" part of the association rule – RHS)	Support	Confidence	Lift	Count
Rule 1					
T5SalesGrowthGT1.05, T5RoICGT12, T33SalesByRcvbl	5YearReturns > NSE 50	0.012	0.78	1.8	933
	3YearReturns > NSE 50	0.012	0.75	1.8	895
Rule 2					
PriceBySalesPerShareLT1, TTMDbyWCLT2, T5RoICGT12, T33CFbyAssets	3YearReturns > NSE 50	0.011	0.75	1.8	848
	5YearReturns NSE 50	0.011	0.72	1.6	807
Rule 3					
T5SalesGrowthGT1.05, TTMDbyWCLT2, T33SalesByRcvbl, T33_T5AvgSalesByT5AvgAssets	5YearReturns > NSE 50	0.013	0.77	1.8	1001
	3YearReturns > NSE 50	0.012	0.7	1.7	913
Rule 4					
T5SalesGrowthGT1.05, T3BVGrowthGT1, T33SalesByRcvbl, T33_T5AvgRoA, T33_T5AvgSalesByT5AvgAssets	5YearReturns > NSE 50	0.011	0.78	1.8	852
	3YearReturns > NSE 50	0.011	0.73	1.8	796
Rule 5					
T5SalesGrowthGT1.05, T3BVGrowthGT1, TTMDERatioLE 1, T33SalesByRcvbl, T33_T5AvgSalesByT5AvgAssets	5YearReturns > NSE 50	0.014	0.78	1.8	1051
	3YearReturns > NSE 50	0.013	0.73	1.8	976
Rule 6					
PriceBySalesPerShareLT1, TTMDbyWCLT2, TTMDERatioLE 1, T5RoICGT12, T33CFbyAssets	3YearReturns > NSE 50	0.011	0.76	1.8	847
	5YearReturns > NSE 50	0.011	0.72	1.7	806
Rule 7					
T5SalesGrowthGT1.05, T3BVGrowthGT1, TTMDbyWCLT2, T33TTM_GrProfitByAssets, T33SalesByRcvbl	5YearReturns > NSE 50	0.011	0.78	1.8	850
	3YearReturns > NSE 50	0.011	0.72	1.7	784

Table 2 (continued)

Antecedent ('If' part of the association rule – LHS)	Consequent ("Then" part of the association rule – RHS)	Support	Confidence	Lift	Count
Rule 8					
T5SalesGrowthGT1.05, T33CFbyAssets, T33SalesByRcvbl, TF_ChgInSalesGTRcvbls, T33_T5AvgSalesByT5AvgAssets	5YearReturns > NSE 50	0.013	0.77	1.8	944
	3YearReturns > NSE 50	0.012	0.7	1.7	863
Rule 9					
T5SalesGrowthGT1.05, TTMDbyWCLT2, T33SalesByRcvbl, T33_T5AvgRoE, T33_T5AvgRoIC	5YearReturns > NSE 50	0.013	0.77	1.8	978
	3YearReturns > NSE 50	0.012	0.7	1.7	897
Rule 10					
T5PATMarginsGT8, T5SalesGrowthGT1, TTMDERatioLE 1, T33TTM_EBITByAssets, T33_T5AvgSalesByT5AvgAssets	5YearReturns > NSE 50	0.012	0.7	1.7	874
	3YearReturns > NSE 50	0.012	0.7	1.8	859

Source: Compiled by the author.

Table 3

Comparison of 'Association-Rule-Portfolio Returns' and 'Index Returns' for the Corresponding Holding Period (Out-of-Sample Data Set)

Performance of 'Rule-1 Portfolio'			
	Holding Period		
	1 Year	**3 Years**	**5 Years**
Np = Number of association rule Portfolios created at different point in time	24	24	24
Number: *Rp* > *Rm* = Number of portfolios with returns > index returns	18	24	24
% *Rp* > *Rm* = Percent of association rule portfolios with returns > index returns	75%	100%	100%
Avg. *Rp*	24.7%	14.8%	16.1%
Avg Mkt Rtn: *Rm*	17.3%	9.9%	10.6%
Performance of 'Rule-2 Portfolio'			
	Holding Period		
	1 Year	**3 Years**	**5 Years**
Np	24	24	24
Number: *Rp* > *Rm*	21	24	23
% *Rp* > *Rm*	88%	100%	96%

Table 3 (continued)

Avg *Rp*	47.7%	34.6%	24.1%
Avg Mkt Rtn: *Rm*	17.3%	9.9%	10.6%
Performance of 'Rule-3 Portfolio'			
	Holding Period		
	1 Year	**3 Years**	**5 Years**
Np	24	24	24
Number: *Rp* > *Rm*	17	24	24
% *Rp* > *Rm*	71%	100%	100%
Avg *Rp*	34.5%	21.2%	19.7%
Avg Mkt Rtn: *Rm*	17.3%	9.9%	10.6%
Performance of 'Rule-4 Portfolio'			
	Holding Period		
	1 Year	**3 Years**	**5 Years**
Np	24	24	24
Number: *Rp* > *Rm*	18	24	24
% *Rp* > *Rm*	75%	100%	100%
Avg *Rp*	34.8%	20.6%	19.9%
Avg Mkt Rtn: *Rm*	17.3%	9.9%	10.6%
Performance of 'Rule-5 Portfolio'			
	Holding Period		
	1 Year	**3 Years**	**5 Years**
Np	24	24	24
Number: *Rp* > *Rm*	19	24	24
% *Rp* > *Rm*	79%	100%	100%
Avg *Rp*	36.6%	22.6%	20.4%
Avg Mkt Rtn: *Rm*	17.3%	9.9%	10.6%
Performance of 'Rule-6 Portfolio'			
	Holding Period		
	1 Year	**3 Years**	**5 Years**
Np	24	24	24
Number: *Rp* > *Rm*	21	24	24
% *Rp* > *Rm*	88%	100%	100%
Avg *Rp*	47.7%	34.6%	24.1%
Avg Mkt Rtn: *Rm*	17.3%	9.9%	10.6%

Table 3 (continued)

Performance of 'Rule-7 Portfolio'			
	Holding Period		
	1 Year	3 Years	5 Years
Np	24	24	24
Number: *Rp* > *Rm*	17	24	24
% *Rp* > *Rm*	71%	100%	100%
Avg *Rp*	30.1%	17.8%	17.3%
Avg Mkt Rtn: *Rm*	17.3%	9.9%	10.6%
Performance of 'Rule-8 Portfolio'			
	Holding Period		
	1 Year	3 Years	5 Years
Np	24	24	24
Number: *Rp* > *Rm*	23	24	24
% *Rp* > *Rm*	96%	100%	100%
Avg *Rp*	40.1%	25.4%	21.3%
Avg Mkt Rtn: *Rm*	17.3%	9.9%	10.6%
Performance of 'Rule-9 Portfolio'			
	Holding Period		
	1 Year	3 Years	5 Years
Np	24	24	24
Number: *Rp* > *Rm*	16	24	24
% Rp > *Rm*	67%	100%	100%
Avg *Rp*	27.4%	17.2%	17.1%
Avg Mkt Rtn: *Rm*	17.3%	9.9%	10.6%
Performance of 'Rule-10 Portfolio'			
	Holding Period		
	1 Year	3 Years	5 Years
Np	24	24	24
Number: *Rp* > *Rm*	17	24	24
% *Rp* > *Rm*	71%	100%	100%
Avg *Rp*	23.2%	13.9%	14.2%
Avg Mkt Rtn: *Rm*	17.3%	9.9%	10.6%