Top performing stocks recommendation strategy for portfolio

Kartikay Gupta\textsuperscript{a*} and Niladri Chatterjee\textsuperscript{a}

\textsuperscript{a}Mathematics Department, IIT Delhi, New Delhi, India

*corresponding author details:

Kartikay Gupta, Mathematics Department, IIT Delhi, New Delhi, Pin:110016, India
e-mail: maz158144@iitd.ac.in.
Top performing stocks recommendation strategy for portfolio

Stock return forecasting is of utmost importance in the business world. This has been the favourite topic of research for many academicians since decades. Recently, regularization techniques have reported to tremendously increase the forecast accuracy of the simple regression model. Still, this model cannot incorporate the effect of things like a major natural disaster, large foreign influence, etc. in its prediction. Such things affect the whole stock market and are very unpredictable. Thus, it is more important to recommend top stocks rather than predicting exact stock returns. The present paper modifies the regression task to output value for each stock which is more suitable for ranking the stocks by expected returns. Two large datasets consisting of altogether 1205 companies listed at Indian exchanges were used for experimentation. Five different metrics were used for evaluating the different models. Results were also analysed subjectively through plots. The results showed the superiority of the proposed techniques.

Keywords: equity returns, forecasting, fundamentals, regression, ranking

1. Introduction

Forecasting of equity price is very important for portfolio management and investment purposes. A lot of literature is available which forecast equity price over different horizons (Rapach, Strauss, and Zhou 2010; Welch and Goyal 2008; Campbell and Thompson 2008; Dangl and Halling 2012; Neely et al. 2014). These researchers have used various indicators including technical and fundamental indicators for the forecasting purpose. The consensus of these studies has been that equity premium is predictable over different horizons. The present work also confirms the above conclusion. It further proposes a novel technique to train parameter for Kitchen sink regression over a big set of fundamental, technical and general indicators. The regression parameters are trained in such a way that output should rank the companies in the order of expected return in next quarter. No importance has been given to
predicting the actual return while training the model. But it has been made sure that a company which gets a higher score on the model is expected to have a higher expected return. This nullifies the effect of unpredictable things like weather conditions or foreign influence on the national economy. This proposed technique will be referred to as rank-regression (RR). The above strategy is successfully able to recommend top companies which give a higher return than the mean performance of all companies. The top 20 companies recommended through the proposed models gave much higher quarterly returns over and above the mean market performance when averaged over large datasets.

The present study takes inspiration from the work of (Li et al. 2017), who statistically showed that equity premium on S&P 500 is predictable through regression with regularisation. The present paper does similar experiments which verify similar results for Indian companies. Secondly, it introduces a novel technique for training the parameter which gives superior results. This novel technique performs better than regularised regression models.

The rest of the paper is organised as follows. Section 2 provides a brief literature review for the subject. Section 3 explains the proposed rank-regression technique and evaluation metrics. Section 4 gives the results of the experiments. Section 5 discusses the finer points of experimental results which demonstrate the superiority of proposed models. Section 6 presents the conclusion.

2. Literature Review

A recent work (Li et al. 2017) in this area has used Kitchen sink regression to achieve this task. They took the forecast horizon of one month and used monthly fundamental indicators. Their model delivered 2.7% excess return over and above other models. While training the parameters for the regression model they made sure that the
coefficients reflect the intuitive relation between the equity fundamentals and the expected return. Secondly, they added L1 and L2 regularisation term to the final error, so that the coefficients remain small. These modifications improved their results tremendously. They took the forecast horizon of 1 month and used monthly fundamental indicators.

(Dangl and Halling 2012) used a stochastic linear regression model to forecast monthly returns of S&P 500 index. They allowed the coefficients of the model to vary over time. The coefficients were modelled as random walk model. The present paper does not allows that kind of uncertainty in the model and rather keeps any uncertainty to be included only in the error term.

(Li, Tsiakas, and Wang 2014) also used kitchen sink regression to forecast exchange rates conditioned on economic fundamentals.

The present paper also utilises a technique built over kitchen sink regression to forecast equity returns conditioned on a big set of important variable. Various research papers are available (Bhar and Nikolova 2009; Tripathi 2008; Bhandari 1988) which analyse the impact of some subset of these variables on equity returns.

The portfolio optimisation problem should be considered as a recommendation problem and not a regression problem. This is because it is easier to compare different companies based on some same underlying features than to forecast their exact individual returns. Thus, the present paper tries to recommend top stocks for any portfolio. Then, those companies should be subjectively scrutinized before purchasing their stocks. (Paranjape-Voditel and Deshpande 2013) used association rule mining to recommend a ranked basket of stocks. They compared their results with the performance of top 5 mutual funds in India.
3. Methodology

The indicators explained in Table 3A (Appendix) were used for forecasting one quarter ahead equity return for each company stock. The target value which was forecasted for each sample is equal to:

\[
\text{Target value} = \frac{P_1 - P_0}{P_0}
\]

where \(P_1\) is one quarter ahead equity price and \(P_0\) is current equity price.

In total 25 indicators were used to train the regression model. The rank-regression model used in the present paper has the following form:

\[
T_j = \sum_{i=1}^{25} p_i \times f_i + e_j
\]

where \(p_i\) is the coefficient of the feature value \(f_i\), \(T_j\) is the target value, and \(e_j\) is the error for the \(j^{th}\) sample.

The missing values in features of the dataset were replaced with zero. Then features were normalized, i.e., the mean and standard deviation of features in training dataset were made equal to 0 and 1 respectively.

The parameters were trained in such a way that model should rank the companies by expected return for each quarter. No incentive was given to model to predict the actual return. This was done in a novel way. Fig 1 shows the pseudo code of the algorithm used to train the parameters of this rank-regression model. Samples in the training data were divided on the basis of quarterly time period. Thus, samples containing every company’s information for a given quarter were all kept in a single batch. Thus, each quarter is represented by a batch which contains all information regarding every company for that period. Forecasts were generated for each batch and compared with the target/actual values to find the loss. The forecasts and the target values were centralized to 0 before calculating the final error. Then, the gradients were
calculated for the loss with respect to each parameter. The parameters were updated based on their gradients to minimize the final error.

**Pseudo Code**

```plaintext
while the stopping criteria is not reached:
    for each batch do:
        \( \alpha_i \) = model prediction of return
        \( \alpha_i \) = actual return
        final prediction for \( i^{th} \) sample (\( \Omega_i \)) = \( \alpha_i \) – mean of \( \alpha \) values over this batch
        actual value for \( i^{th} \) sample (\( \Omega_i \)) = \( \alpha_i \) – mean of \( \alpha \) values over this batch
        final error = \( \Sigma_i (\Omega_i - \Omega_i)^2 \)
    perform Gradient Descent over the model parameters to minimize the final error
```

**Figure 1**: Proposed Algorithm, Pseudo code for the proposed algorithm used to train the model parameters through Gradient Descent (GD).

Here batch centralization was done to tide over things which affect the whole stock market. Such things like foreign influence, major revisions in government policies, weather etc affect the whole market as a whole, and their influence needs to be zeroed off. Thus, training was done batch wise with centralisation as explained earlier. The model should be able to rank the companies based on expected return for a desired quarter. While testing the model the output values were just used to rank the companies for each quarter. Finally, top companies based on this ranking were chosen for the portfolio. Figure 1 gives the pseudo-code for the proposed algorithm.

(Li et al. 2017) compared the performance of regression models with different kind of regularisation (L1 regularisation, L2 regularisation or both). They reported that Ridge KS regression, which uses L2 regularisation, performs best amongst these 3 models during the period of expansion. Expansion denotes the period of the economy
where stock prices generally grow higher, as defined by National Bureau of Economic Research (NBER), USA. Though no such officially defined cycles exist for the Indian economy, but the data set in the present paper can be classified into expansion, as the stock prices steadily increased over these years.

Owing to the reported success of regularisation in such task (Li et al. 2017), L2 regularisation was done to the trainable parameters while being trained through Stochastic gradient descent. This means sum of squares of the coefficients was added to the final loss function. Thus, the coefficients were trained with the dual objective of obtaining correct forecasts and keeping the coefficient size small. The results obtained through the proposed rank-regression model with regularisation are reported separately from the results of the rank-regression model. Thus, the proposed models were compared with a simple regression model, a ridge regression model and a naïve model. The simple regression model and ridge regression model used the same input features to train a linear relationship between the inputs and the quarterly returns, as in the proposed models. These models differed only in the way the parameters were learned through training on the train set. The naïve model simply ranked the companies based on the last quarter returns. The best-performing companies in the last quarter are expected to perform best in the current quarter as well. This is inspired by the Wiener process or random walk based modelling of stock prices. Thus, the experiments were performed on these five models to ascertain the best models amongst these. Table 1 succinctly describes the models, leaving out the naïve model. The results verified the superiority of the proposed technique.
Table 1: Model descriptions of different models. Feature normalization refers to normalization of input features. Target value normalization refers to normalization of target value using mean and standard deviation of the training data test values. Thus, training data target values mean and standard deviation becomes equal to 0 and 1 respectively. But, test data target values mean and standard deviation is not exactly equal to 0 and 1 respectively. ‘\( \bar{o} \)’ and ‘\( \bar{a} \)’ refers to mean of model output values and actual values for a batch respectively. ‘\( p \)’ refers to regression equation coefficient value.

| Model       | Pre-processing steps                  | Final error function                                      | Training Procedure                                                                 |
|-------------|---------------------------------------|-----------------------------------------------------------|-----------------------------------------------------------------------------------|
| 1           | Simple SGD                            | Feature Normalisation + Target value Normalisation        | Parameters are updated through SGD after each iteration consisting of only 1 sample.|
| 2           | Simple SGD + L2                       | Feature Normalisation + Target value Normalisation        | Parameters are updated through SGD after each iteration consisting of only 1 sample.|
| 3           | Rank-regression                       | Feature Normalisation                                     | Parameters are updated through SGD after each iteration consisting of all samples within each quarter.|
| 4           | Rank-regression + L2                  | Feature Normalisation                                     | Parameters are updated through SGD after each iteration consisting of all samples within each quarter.|

3.1 Evaluation metrics

The models mentioned above were used to rank the stocks by expected next quarter return. The models were compared on the basis of three metrics:

1.) **AP@100**: Average Precision Metric for top 100 recommendations (AP@100).

This measure is given by:

\[
AP@100 = \frac{1}{100} \times \sum_{k=1}^{100} P(k).rel(k)
\]  

(3)

where P(k) is obtained after dividing ‘number of correct recommendations amongst top k companies’ by ‘k’. ‘rel(k)’ is an indicator function which is 1 when k\(^{th}\) recommendation is amongst top k and otherwise 0.

2.) **Top 20**: Actual Return generated by investing equally in top 20 stocks as recommended by a model

3.) **Top 50**: Actual Return generated by investing equally in top 50 stocks as recommended by a model
4.) **Riskless 20**: Non-dominated solutions with low risk and high return were calculated. Riskless 20 gives mean-return of top 20 stocks lying on the front of least risk and highest return. The standard deviation of past returns was used as a risk measure. Within each front, stocks were sorted according to higher return.

5.) **Riskless 50**: Similar to Riskless 20, Riskless 50 gives mean-return of top 50 stocks lying on the front of least risk and highest return.

The present paper uses a simple portfolio investment strategy of investing equally in all the top recommended stocks. Mean-variance trading strategy has been used in past researches (Li et al. 2017) to create a portfolio based on expected return and variance of each stock. But the present paper cannot use that strategy as expected return forecasts are not generated by the proposed model. It recommends top companies and then invests equally in top few companies.

**4. Results**

Two datasets were extracted from Thomson Reuters Eikon tool to perform the experiments. The first data set consisted of 497 companies listed at BSE 500 index. The second data set consisted of 708 companies traded at National Stock Exchange (NSE) of India. Those companies were removed from the dataset which had a high number of unavailable values for the mentioned indicators during any of the quarters.

The companies’ information were extracted quarterly for a period of seven years. The period considered started from the January-march quarter of 2011 and ended at october-december quarter of 2017. Thus in total, the data sets contained information and a target value for companies for these 28 quarters.
In each experiment, a neural network was trained on past quarter’s information and was used to rank all the listed stocks in the next quarter. An expected relative return was generated by the neural network for each stock used for training. This relative return finally determined the rank or performance of that stock amongst all others. Thus 25 parameters need to be learnt for each of the dataset consisting of either 497 companies or 708 companies. This means that these parameters should minimise the error in all of these 497 or 708 companies simultaneously. This had to be done for each of the training set of the train-test pairs. Therefore, the dataset is large and requires huge computation for calculating parameters. Thus stochastic gradient descent methodology was used for parameters training. This is the same methodology which is used for learning deep neural networks, where it is more sophisticatedly referred to as back propagation algorithm.

An error was calculated for each stock, and the parameters were updated to minimise that error. This was done till the stopping criteria was reached. The stopping criteria stopped the iteration procedure whenever the train had reached its minimum value while the training had not improved since last 30,000 iterations. The maximum number of iterations were capped at 8,00,000 though iterations stopped mostly before that number.

Comparative experiments were conducted on the different models through extended cross-validation. The cross-validation was done through organising data into rolling windows, where training was done on a certain number of consecutive quarters and the next quarter was used for testing purpose. The number of consecutive quarters used for training in 3 different experiments were 10, 15 and 20. Thus in total results were validated on 18, 13 and 8 pairs of train-test data sets corresponding to experiments with 10, 15 and 20 training-timestamps/training-quarters respectively.
The algorithms were coded into python 3 for experimentation. Pytorch, a python deep learning module, was utilised for learning the regression parameters.

The results for all the three experiments demonstrate the superiority of the proposed models. Table 1,2 show results when 10 quarters used for learning the parameters. Similarly, Table 3,4 correspond to 15 quarters and table 5,6 correspond to 20 quarters.

**Table 1:** Results corresponding to BSE Data set and 10 number of training quarters.

| Model Name | Evaluation Metrics |
|------------|--------------------|
|            | AP@100 | Top 20 | Top 50 | Riskless 20 | Riskless 50 |
| Rank-Regression (RR) + L2 | **0.03256** | **0.16751** | **0.14684** | 0.11672 | **0.11245** |
| RR         | 0.02736 | 0.12361 | 0.12747 | 0.10809 | 0.09784 |
| Ridge      | 0.03172 | 0.16544 | 0.14177 | **0.12192** | 0.10888 |
| Simple SGD | 0.02914 | 0.14011 | 0.13848 | 0.10535 | 0.10194 |
| Naïve      | 0.01303 | 0.08533 | 0.07882 | 0.06308 | 0.06490 |

**Table 2:** Results corresponding to NSE Data set and 10 number of training quarters.

| Model Name | Evaluation Metrics |
|------------|--------------------|
|            | AP@100 | Top 20 | Top 50 | Riskless 20 | Riskless 50 |
| RR + L2    | **0.01590** | 0.13808 | 0.13250 | **0.14037** | **0.11722** |
| RR         | 0.01492 | 0.13885 | 0.12970 | 0.11910 | 0.09885 |
| Ridge      | 0.01547 | **0.13974** | 0.13435 | 0.14015 | 0.11270 |
| Simple SGD | 0.01522 | 0.13606 | **0.13686** | 0.13632 | 0.10873 |
| Naïve      | 0.00513 | 0.08994 | 0.07334 | 0.05761 | 0.07052 |
Table 3: Results corresponding to BSE Data set and 15 number of training quarters.

| Model Name   | AP@100  | Top 20 | Top 50 | Riskless 20 | Riskless 50 |
|--------------|---------|--------|--------|-------------|-------------|
| RR + L2      | 0.03203 | 0.11631| 0.10228| 0.07355     | 0.08015     |
| RR           | 0.03183 | 0.11644| 0.09927| 0.08656     | 0.07329     |
| Ridge        | 0.03191 | 0.12486| 0.09808| 0.07647     | 0.07915     |
| Simple SGD   | 0.02888 | 0.11669| 0.09660| 0.07924     | 0.07772     |
| Naïve        | 0.01300 | 0.04686| 0.04191| 0.04135     | 0.03756     |

Table 4: Results corresponding to NSE Data set and 15 number of training quarters.

| Model Name   | AP@100  | Top 20 | Top 50 | Riskless 20 | Riskless 50 |
|--------------|---------|--------|--------|-------------|-------------|
| RR + L2      | 0.01643 | 0.09844| 0.10033| 0.09247     | 0.07627     |
| RR           | 0.01882 | 0.11941| 0.10646| 0.09991     | 0.08249     |
| Ridge        | 0.01665 | 0.09839| 0.09586| 0.09363     | 0.07869     |
| Simple SGD   | 0.01704 | 0.10528| 0.10235| 0.07885     | 0.05952     |
| Naïve        | 0.00689 | 0.06343| 0.05698| 0.02914     | 0.04165     |

Table 5: Results corresponding to BSE Data set and 20 number of training quarters.

| Model Name   | AP@100  | Top 20 | Top 50 | Riskless 20 | Riskless 50 |
|--------------|---------|--------|--------|-------------|-------------|
| RR + L2      | 0.03276 | 0.15115| 0.12630| 0.10537     | 0.10268     |
| RR           | 0.03219 | 0.13716| 0.12328| 0.10072     | 0.09683     |
| Ridge        | 0.03134 | 0.15018| 0.12572| 0.10661     | 0.10162     |
| Simple SGD   | 0.02720 | 0.12734| 0.12326| 0.10460     | 0.09768     |
| Naïve        | 0.01107 | 0.08106| 0.06676| 0.06829     | 0.05948     |
Table 6: Results corresponding to NSE Data set and 20 number of training quarters.

| Model Name | AP@100 | Top 20 | Top 50 | Riskless 20 | Riskless 50 |
|------------|--------|--------|--------|-------------|-------------|
| RR + L2    | 0.01753| 0.11727| 0.12319| 0.10021     | 0.10575     |
| RR         | 0.02116| **0.12968**| 0.11513| 0.08292     | 0.08120     |
| Ridge      | 0.01782| 0.12102| **0.12389**| **0.11436**| **0.10833**|
| Simple SGD | 0.02099| 0.12514| 0.11699| 0.09418     | 0.08668     |
| Naïve      | 0.00650| 0.06569| 0.06514| 0.04615     | 0.05956     |

The metric AP@100 is consistently higher in all the three experiments for the proposed models. The results for NSE data set on 15 quarters particularly stand out because here the rank-regression model without regularisation outperforms all the rest models in all the evaluation metrics.

In all the previous six tables the proposed models outperform the existing models in at least three of the evaluation metrics except one of the tables. In that exceptional table, corresponding to NSE data set trained on 20 quarters, still AP@100 and Top 20 returns metrics point towards the superiority of the rank-regression models.

Further analysis was done on results obtained by training the model on 26 quarters and testing on last two quarters. Here the results were subjectively analysed through figures, which solidify the significance of the proposed models. The numerical results corresponding to this experiment are given in Table 1A and 2A (Appendix)
Figure 2: Cumulative Mean Returns. Mean return of companies ranked according to expected return as per each of the models. The number of companies plotted for each model has been truncated at 100. Each sub-figure denotes different dataset or period, i.e., (i) Q3 2017-18, BSE dataset, (ii) Q4 2017-18, NSE dataset, (iii) Q4 2017-18, BSE dataset, (i) Q3 2017-18, NSE dataset.

Figures 2 and 1A (Appendix) give the plot of cumulative mean-return for companies ranked according to higher output in each model. The steps to obtain each plot are:

1. For each model, we obtain the ranks for each company by sorting as per the expected relative return generated by the model.
2. Then we take the mean of actual returns for top ‘n’ companies.
3. This gives us the value of plot at value ‘n’ on x-axis.
4. The y-axis values correspond to the returns generated when an equal amount of money would have been invested in each of the top ‘n’ companies.
In figures 2 and 1A (Appendix), it can be seen that the proposed models’ plot are mostly higher than the other models’ plot. The rank-regression models’ plot is consistently higher than others at least for the top 100 companies. This means that equal investment in top companies recommended by the proposed models generates more profit than investing in companies recommended by other models. The cumulative mean return remains high for the initial few companies and then gently slopes down to the lowest value at the end. This curve perfectly depicts that the model has been able to rank the companies correctly. Companies having higher rank do tend to show a higher return in that quarter. After 100 companies, mean cumulative return becomes more or less same for all the models. Thus, the proposed model has been successfully able to single out top performers amongst different stocks.

Ensemble models are predicting models which combine forecasts from two or more models to get final prediction value. They generally perform better than individual models and there exist theoretical and empirical foundations (Hagedorn, Doblas-Reyes, and Palmer 2005) for this result. (Zhang 2003; Firmino, de Mattos Neto, and Ferreira 2015; Khashei and Bijari 2011) used hybrid models to achieve better performance for prediction. In an extension of this work, the proposed model may be combined with other techniques to achieve superior results.

5. Conclusion

The present paper proposes a novel technique for recommending top stocks in any period. Two large data sets consisting of 497 companies listed at BSE (India) and 708 companies listed at NSE (India), were used for the experiments. The results were evaluated using five evaluation metrics. The results proved the superiority of the proposed technique. Results were also subjectively analysed through figures, which
clearly showed the effectiveness of the rank-regression. The research paper also verified that L2 regularization is very effective in improving equity forecast results for the Indian stock market, as reported in past studies for other markets.

Acknowledgements
This work was supported by the Ministry of Human Resource Development (MHRD) through the Teaching Assistantship Grant at IIT Delhi, India.

References

Bhandari, Laxmi Chand. 1988. “Debt/Equity Ratio and Expected Common Stock Returns: Empirical Evidence.” *The Journal of Finance* 43 (2). Wiley/Blackwell (10.1111): 507–28. https://doi.org/10.1111/j.1540-6261.1988.tb03952.x.

Bhar, Ramaprasad, and Biljana Nikolova. 2009. “Oil Prices and Equity Returns in the BRIC Countries.” *World Economy* 32 (7). Wiley/Blackwell (10.1111): 1036–54. https://doi.org/10.1111/j.1467-9701.2009.01194.x.

Campbell, John Y., and Samuel B. Thompson. 2008. “Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?” *Review of Financial Studies* 21 (4): 1509–31. https://doi.org/10.1093/rfs/hhm055.

Dangl, Thomas, and Michael Halling. 2012. “Predictive Regressions with Time-Varying Coefficients.” *Journal of Financial Economics* 106 (1). North-Holland: 157–81. https://doi.org/10.1016/J.JFINECO.2012.04.003.

Firmino, Paulo Renato A., Paulo S.G. de Mattos Neto, and Tiago A.E. Ferreira. 2015. “Error Modeling Approach to Improve Time Series Forecasters.” *Neurocomputing* 153: 242–54. https://doi.org/10.1016/j.neucom.2014.11.030.

Khashei, Mehdi, and Mehdi Bijari. 2011. “A Novel Hybridization of Artificial Neural Networks and ARIMA Models for Time Series Forecasting.” *Applied Soft Computing* 11 (2): 2664–75. https://doi.org/10.1016/j.asoc.2010.10.015.

Li, Jiahan, Ilias Tsiakas, Jiahan Li, and Ilias Tsiakas. 2017. “Equity Premium Prediction: The Role of Economic and Statistical Constraints.” *Journal of Financial Markets* 36. Elsevier: 56–75. https://econpapers.repec.org/article/eefinmar/v_3a36_3ay_3a2017_3ai_3ac_3ap_3a56-75.htm.

Li, Jiahan, Ilias Tsiakas, and Wei Wang. 2014. “Predicting Exchange Rates Out of Sample: Can Economic Fundamentals Beat the Random Walk?” *E Journal of
Financial Econometrics.
http://www.uoguelph.ca/~itsiakas/papers/FX_Fundamentals.pdf.

Neely, Christopher J., David E. Rapach, Jun Tu, and Guofu Zhou. 2014. “Forecasting the Equity Risk Premium: The Role of Technical Indicators.” Management Science 60 (7): 1772–91. https://doi.org/10.1287/mnsc.2013.1838.

Paranjape-Voditel, Preeti, and Umesh Deshpande. 2013. “A Stock Market Portfolio Recommender System Based on Association Rule Mining.” Applied Soft Computing 13 (2). Elsevier: 1055–63. https://doi.org/10.1016/J.ASOC.2012.09.012.

Rapach, David E., Jack K. Strauss, and Guofu Zhou. 2010. “Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy.” Review of Financial Studies 23 (2): 821–62. https://doi.org/10.1093/rfs/hhp063.

Tripathi, Vanita. 2008. “Company Fundamentals and Equity Returns in India.” SSRN Electronic Journal, May. https://doi.org/10.2139/ssrn.1134651.

Welch, Ivo, and Amit Goyal. 2008. “A Comprehensive Look at The Empirical Performance of Equity Premium Prediction.” Review of Financial Studies 21 (4): 1455–1508. https://doi.org/10.1093/rfs/hhm014.

Zhang, G.Peter. 2003. “Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model.” Neurocomputing 50: 159–75. https://doi.org/10.1016/S0925-2312(01)00702-0.
Appendix

Figure 1A: Cumulative Mean Returns, Mean return of companies ranked according to expected return as per each of the models. Each sub-figure denotes different dataset or period, i.e., (i) Q3 2017-18, BSE dataset, (ii) Q4 2017-18, NSE dataset, (iii) Q4 2017-18, BSE dataset, (i) Q3 2017-18, NSE dataset

Table 1A: gives evaluation scores for BSE data set.

| Model Name | AP@100 | Top 20 | Top 50 |
|------------|--------|--------|--------|
| RR + L2    | 0.0354 | 0.1599 | 0.1009 |
| RR         | 0.0256 | 0.1124 | 0.0985 |
| Ridge      | 0.0390 | 0.1118 | 0.1095 |
| Simple SGD | 0.0325 | 0.0962 | 0.0944 |
| Naïve      | 0.0278 | 0.1370 | 0.0944 |
Table 2A: Evaluation scores for NSE data set.

| Model Name | Evaluation Metrics |
|------------|--------------------|
|            | AP@100             | Top 20 | Top 50 |
| RR + L2    | 0.0185             | 0.1582 | 0.0902 |
| RR         | 0.0256             | 0.1541 | 0.0898 |
| Ridge      | 0.0171             | 0.0991 | 0.0969 |
| Simple SGD | 0.0157             | 0.0661 | 0.0693 |
| Naïve      | 0.0124             | 0.0005 | 0.0387 |

Table 1A and 2A give the evaluation metric results for the five different models for BSE and NSE dataset respectively. The mean quarterly return of all the companies for BSE and NSE dataset was 0.0327 and 0.05 respectively, which is quite less than the returns obtained through the proposed model. The top scores in each column are bolder than the others. The rank-regression models consistently perform better than the other models. The quarterly return on top 20 companies recommended by the proposed model with regularisation is far better than any other model. It defeats other models by at least 2.29% at quarterly returns in both the datasets. The other metrics also show that the proposed models give optimum results. The proposed models’ performance is relatively more superior in NSE data set comprising of 708 companies which is much larger than the BSE dataset (497 companies). Even in BSE data set, the proposed models’ performance is higher or almost optimum for each of the three metrics.
| NAME                                      | TYPE                        | DESCRIPTION                                                                                                                                 |
|-------------------------------------------|-----------------------------|-------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | Gross Profit Margin Industrial And Utility Pct | Company Fundamental        | This item is calculated as Gross Profit (Industrial and Utility) for the fiscal period divided by Primary Revenue for the same period, and multiplied by 100. |
| 2  | Oper Profit Margin Pct                   | Company Fundamental        | This value measures the percent of revenues remaining after paying all operating expenses. It is calculated as Operating Income divided by Total Revenue for the fiscal period, multiplied by 100. |
| 3  | Income Aft Tax Margin pct                | Company Fundamental        | Income after Tax Margin is also known as Return on Sales. This value is the Income After Taxes for the fiscal period divided by Total Revenue for the same period and is expressed as a percentage. |
| 4  | Asset Turnover                           | Company Fundamental        | The amount of revenue generated for each unit of assets. Also known as TAT. It is calculated as Primary Revenue for the fiscal period divided by the Average Total Assets for the same period. |
| 5  | Return on Avg To Eqty Pct Net Income Before Exa Items | Company Fundamental        | This value is calculated as the Net Income Before Extraordinary Items for the fiscal period divided by the same period Average Total Equity and is expressed as a percentage. Average Total Equity is the average of Total Equity at the beginning and the end of the year. |
| 6  | Hist To tDebt Comm Eqty Pct              | Company Fundamental        | This is the percentage of Total Debt as of the end of the fiscal period to Common Shareholders Equity for the same period. Not available for Banks. |
| 7  | Company Market Cap                       | Company Fundamental        | The Company Market Capitalization represents the sum of market value for all relevant issue level share types. The issue level market value is calculated by multiplying the requested shares type by latest close price. |
| 8  | PE                                       | Company Fundamental        | A valuation ratio of a company’s current share price relative to its EPS. EPS is LTM Earnings per Share From Continuing Operations. PE is not calculated when LTM EPS is less than or equal to Zero. |
| 9  | EV To EBITDA                             | Company Fundamental        | This ratio measures how much a company is valued per each dollar of EBITDA. EBITDA is LTM Earnings before Interest, Taxes, Depreciation and Amortization. EV represents the sum of Market Capitalization, Total Debt, Preferred Stock and Minority Interest minus Cash and Short Term Investments for the most recent fiscal period. |
| 10 | Dividend Yield                           | Company Fundamental        | The ratio of the annualised dividends to the price of a stock. Dividends are adjusted to account for any stock splits during the 12-month period. |
| 11 | EV To Sales                              | Company Fundamental        | EV represents the sum of Market Capitalization, Total Debt, Preferred Stock and Minority Interest minus Cash and Short Term Investments for the most recent fiscal period. Sales is LTM Total Revenue. |
| 12 | Price To CF Per Share                    | Company Fundamental        | This is the Current Market Capitalization divided by the latest annual Cash Flow. Cash Flow represents Net Income After Taxes minus Preferred Dividends and General Partner Distributions plus Depreciation and Amortization of Intangibles. |
| 13 | Price To BV Per Share                    | Company Fundamental        | Price To Book Value Per Share is calculated by dividing the company’s latest closing Price by its Book |
Value per share. Book Value per share is calculated by dividing Total Equity from latest fiscal period by Current Total Shares Outstanding.

|   |   | Company Technicals |   |
|---|---|---------------------|---|
| 14 | avg Q rtn n | Average return for each quarter since last n quarters. This is calculated by dividing total return in last n quarters by n and the stock price before n quarters. The following indicators were used: avg Q rtn 8, avg Q rtn 7, avg Q rtn 6, avg Q rtn 5, avg Q rtn 4, avg Q rtn 3, avg Q rtn 2, avg Q rtn 1. |   |
| 15 | USD/INR | Economical | US Dollar/ Indian Ruppee spot rate |
| 16 | IN10YT=RR | Economical | India government 10-yr bond yield |
| 17 | INRPM=RBI A | Economical | India Repo Rate Liquidity Adjustment Facility |
| 18 | MCGBc1 | Economical | Multi Commodity Exchange Of India Crude Oil Energy Future |