Special Proceeding of the 22nd Annual Conference of SSCA
Savitribai Phule Pune University, Pune, January 02-04, 2020; pp 73–84

Index Tracking for NIFTY50

Divyashish Choudhary[1] and Rituparna Sen[2]
[1]*Indian Institute of Technology Madras, Chennai 600036, India*
[2]*Indian Statistical Institute, Bengaluru 560059, India*

Received: 28 May 2020 ; Revised: 25 June 2020 ; Accepted: 27 June 2020

Abstract

Passive investing has been on the rise in global markets for the last four decades now. It constitutes 45% of U.S. stock based funds up from 25% in just a decade. The prominent reason is the exorbitant costs of active management which are more often than not unjustified by their performance in relation to benchmark index. Passive management mitigates this issue by closely tracking the benchmark index with minimum transaction costs and management fees.

Although there has been expansive research on passive investing in developed markets like the U.S. capital market, the issue has been covered in marginal detail in emerging markets like that of India. Index tracking is at the heart of passive investing and this paper aims to discern the efficacy of our method of construction of an index tracker in the Indian market while focusing on NIFTY50 index. We employed lowess smoothing method and subsequently the partial correlation to create a tracker with subset of the 50 stocks that constitute the benchmark. Further, we quantified the effects that changing rebalancing frequency and number of constituent stocks in the tracker had on the tracking error and transaction costs and suggested optimal trackers.

Key words: Passive Investing; Lowess Smoothing; Transaction Costs; Tracking Error.

AMS Subject Classifications: 60G50, 05C81

1. Introduction

Active investing has been around since the inception of modern money and capital markets. It is build on the philosophy that one can get better returns than market on an average if the organisation or person allocating the capital is skilled enough to exploit the inefficiencies which are assumed to be present in the market.

Passive investing on the other hand assumes that markets are efficient and the best returns that one can get are by investing capital in the portfolio which mimics market composition. This philosophy has gained prominence mainly because majority of active portfolio managers have a track record that trails market returns when taken over a sufficiently large period like 10-20 yrs. This empirical evidence coupled with the fact that passive investing

Correponding Authors: Divyashish Choudhary, Rituparna Sen
Email: divyashishchoudhary@gmail.com and rsen@isibang.ac.in

is simple and has lower transaction charges has led to capital's exodus from active portfolio management into passive portfolio management. Although passive investing can have many forms, index investing is the most common. It entails building a portfolio which mimics a major benchmark index in the market and hence gives an investor returns in line with the broader market which is the essence of passive investment. This strategy has minimum buy/sell operations and hence reduces market friction which reduces the returns of an investor over the long term.

There are various methods to track an index, simplest being full replication of an index. This method gives minimum tracking error but incurs explicit and implicit trading costs which makes it the costliest strategy. Second method can be optimization based on mean/variance analysis (Roll (1992)). This is a solution which holds fewer stocks than full replication but still doesn't compromise much on the tracking error. However, studies (Focardi and Fabozzi (2004)) note that the noise can dominate the correlations of stocks which renders variance/covariance unreliable. This stems from the fact that high dimensional covariance matrices cannot be estimated consistently. Nakayama and Yokouchi (2018) propose a method that picks constituent stocks for a tracker based on the similarities they have to benchmark index. In turn, these similarities are arrived at by calculating distances of time series trends which are derived from decomposing original series using lowess smoothing.This does not yield appreciable results due to two factors. The first being the residual time series left after lowess decomposition and second being the high correlation between the stocks that constituted the tracker. To improve tracking performance, they propose a similarity-balanced approach in which different groups are formed based on ranks generated from the similarity approach and representative stocks are taken from each group.

We have built upon this method by utilizing iterative confounding variable approach to overcome the correlation issue instead of similarity-balanced approach. This method works in this case because the number of data points far exceed the number of constituent stocks and hence the variance-covariance matrix has eigenvalues different from those stipulated by random matrix theory distribution (Focardi and Fabozzi (2004)). Unfortunately, we were not able to find any study of similar kind done for the Indian market which left us with no reference or comparison source for the results we got.

This paper proceeds as follows. Section 2 describes the data used for the study. Section 3 expounds the methodology starting with lowess smoothening continuing on to the use of partial correlation and how the transaction costs were calculated. Section 4 presents the results we obtained for our index-tracking portfolio , the transaction costs incurred and the best optimal combination of rebalancing window and number of constituent stocks for our tracker. Section 6 concludes and presents scope for further research.

2. Data

Daily closing price adjusted for dividend, bonus and splits has been taken with starting date at 01/06/2008 and ending date at 01/06/2018 for the 50 constituent stocks and the NIFTY50 index. The data was retrieved from yahoo finance website. Below is the snippet of the raw data and processed data in which we combined the adjusted close data of all the stocks and discarded other data columns.

Date	Open	High	Low	Close	Adj Close	Volume
02/06/2008	159.100006	165.399994	152.009995	153.160004	135.345947	4245850
03/06/2008	148.199997	153.979996	146.669998	147.880005	130.680008	2664250
04/06/2008	148.399994	151.139999	138.449997	140.440002	124.105370	1915440
05/06/2008	140.199997	141.779999	127.019997	131.509995	116.214043	3579815
06/06/2008	133.979996	136.250000	127.000000	127.919998	113.041603	2841970
09/06/2008	124.000000	125.199997	118.260002	122.519997	108.269669	2306670
10/06/2008	119.839996	123.769997	116.000000	117.849998	104.142845	2586135
11/06/2008	119.599998	126.279999	119.040001	121.419998	107.297615	3841800
12/06/2008	119.580002	120.879997	114.000000	118.699997	104.893974	2530485

Date	Price_1	Price_2	Price_3	Price_4	Price_5	Price_7	Price_8
2008-06-02	153.160004	91.800003	126.129997	29.375000	149.869995	25.003599	58.099998
2008-06-03	147.880005	92.400002	124.375000	28.900000	151.479996	23.760000	58.633301
2008-06-04	140.440002	87.599998	125.000000	27.910000	147.360001	22.754400	53.716702
2008-06-05	131.509995	85.800003	125.665001	28.665001	154.179993	22.322100	50.308300
2008-06-06	127.919998	85.400002	126.080002	28.230000	152.110001	21.510799	50.158298
2008-06-09	122.519997	81.900002	125.605003	27.165001	140.619995	18.610600	46.474998
2008-06-10	117.849998	82.349998	125.665001	28.090000	136.300003	18.596100	47.433300
2008-06-11	121.419998	88.250000	128.785004	30.094999	141.160004	17.337900	45.541698
2008-06-12	118.699997	85.900002	125.904999	31.410000	142.869995	16.385700	44.891701

Figure 1: Left: raw data for a stock ; Right: pre-processed data for all stocks

3. Methodology

3.1. Lowess smoothing of the raw data

We have taken the data of all the stocks and the index in the time frame marked out in section 2. We have applied *Lowess* smoothing as proposed by Cleveland (1979); used by Shibata and Miura (1997) as well as Nakayama and Yokouchi (2018). *Lowess* smoothes values by applying locally weighted linear regression with weights determined by the distance of data points from the point selected for smoothing. It is a non-parametric method of smoothing the data which assumes no prior assumptions about economic cycles or specific models.

We have decomposed the price time series data into components of long term trends and short term trends. The long term trend was obtained by smoothing the normalised price time series of all the constituents and the index. Short term trends were then obtained by smoothing residuals from the long term smoothing. The residual after second smoothing was weakly-stationary. We have kept one year as the time frame for long term trends and 1 month as the time frame for short term trends. Figure 1 illustrates such break-up for the NIFTY50 index.

3.2. Tracker construction using partial correlation

Nakayama and Yokouchi (2018) used similarity between constituent stocks and the index to rank stocks; out of which the top few were selected to construct a tracker. This was unlike the previous studies which used similarity between the constituent stocks to form clusters (Focardi and Fabozzi (2004), Dose and Cincotti (2005)) or which used integration (Thomaidis (2013), Papantonis (2016)) for index tracking. However, they were not able to beat the results obtained by clustering techniques through the similarity approach. Hence, they developed similarity-balanced approach to tackle the issue of correlation between the constituent stocks and now were able to provide better returns than the similarity approach.

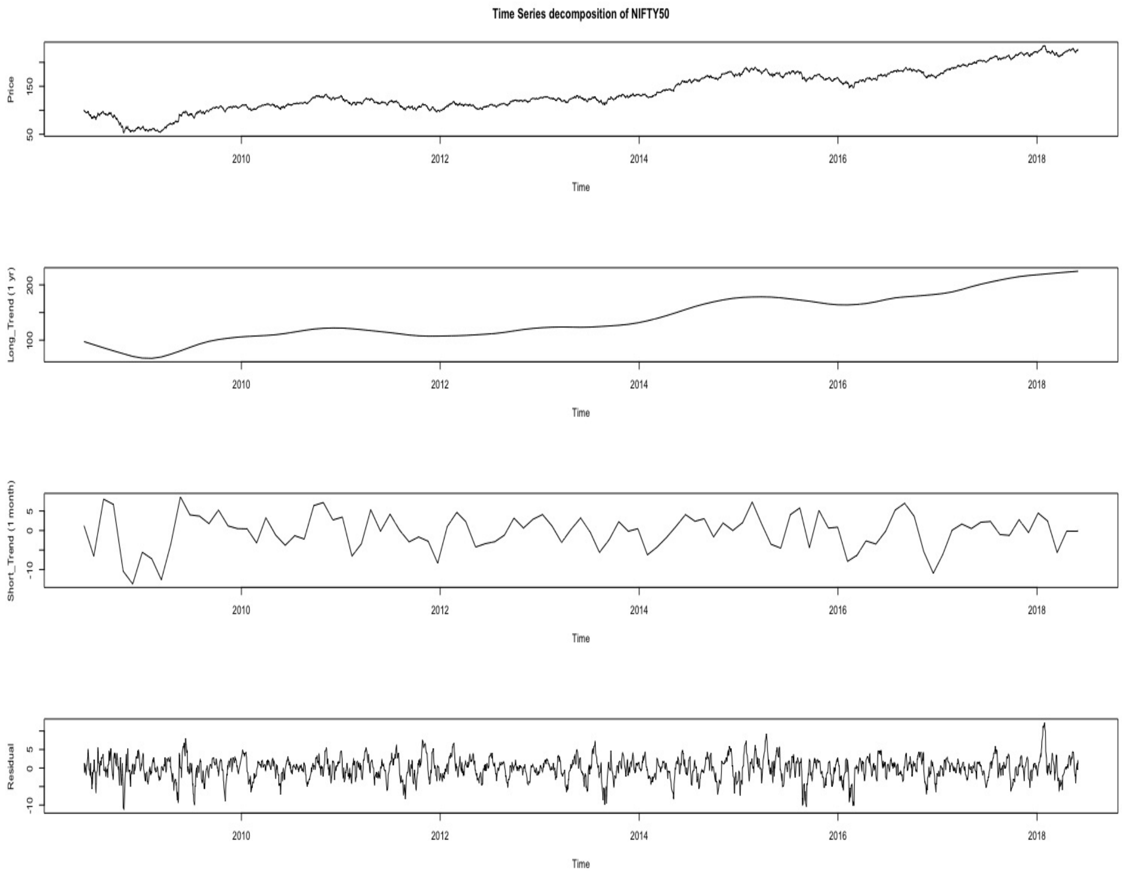

Figure 2: Lowess smoothing and subsequent decomposition of NIFTY50

In our paper, we have incorporated the operational features of similarity-balanced approach by computing inter-day percentage change data for both long term and short term trends and then using partial correlation between the percentage change data of index and the constituent stocks to rank the stocks and construct an index based on this ranking. We implemented this procedure separately between long term trends of stocks & the index and between short term trends of stocks & the index. Two different rankings and subsequently two different trackers were obtained using this methodology. This technique allowed us to eliminate confounding bias but in a statistical fashion rather than heuristic fashion.

We have assigned equal weight to all the constituents of a tracker. On rebalancing, the partial correlation is again calculated in a similar manner as stated before and some constituents are replaced but the total capital at that point is again redistributed equally among all the new constituents and old constituents which still remain. Finally, we have compared and suggested winning tracker on the basis of reduction of tracking error and transaction costs. Tracking error in our context is defined as the difference in percentage change between benchmark index and the tracker after every rebalancing window. Transaction cost's definition and calculation will be detailed in the next sub-section. We have used three parameters to control the values of costs and tracking error - number of stocks in the tracker, rebalancing frequency and composition of hybrid long-short tracker.

3.3. Transaction costs

Transaction costs play the biggest role in determining the real returns that an index can provide. We aim to minimize the transaction costs which we will assume to have both a proportional and fixed component for every trade (Jha and Srivastava (2013), Kellerer et.al (2000), Nakayama and Yokouchi (2018)). Let p denote the per rupee transaction cost, f denote the fixed cost per transaction, n denote the number of stocks in the tracker, R_i denote the return on stock i (value at end of rebalancing period/value at the beginning of it), MAD be mean absolute deviation, turnover be the ratio of number of stocks replaced at rebalancing point to the total number of stocks in the tracker, j be the set of stocks that will be replaced at rebalancing point and l be the total number of these replaced stocks. Then the proportional and fixed costs will be given as follows with detailed derivation attached in the appendix.

$$Cost_{proportional} = pMAD(R_i) + 2p(turnover)[\frac{1}{l}\sum_j min(R_j, \bar{R})]$$

$$Cost_{fixed} = 2fl + f(n-l)$$

4. Results

4.1. Number of stocks in tracker

We have taken 2 years as the training period for construction of all the trackers in this and the ensuing sub-sections. Tables 1 and 2 show the results obtained when we varied the number of stocks in the 2 different categories of trackers - obtained by using short trends and long trends respectively with Quarterly rebalancing window.

Table 1: Effect of no. of stocks on short-trend tracker

No. of Stocks	Average Turnover	Average Rebalancing Cost	Iter. $\leq 1\%$ Tracking error	Iter. $\geq 4\%$ Tracking error	OU Ratio
5	0.29	0.25%	31%	38%	1.67
10	0.28	0.37%	28%	28%	1.46
15	0.26	0.49%	19%	34%	3
20	0.22	0.58%	22%	28%	3
25	0.18	0.67%	16%	28%	3
30	0.14	0.75%	16%	31%	2.56
35	0.09	0.82%	31%	09%	0.6
40	0.05	0.87%	16%	25%	3
45	0.01	0.93%	16%	19%	2.56

Table 2: Effect of no. of stocks on long-trend tracker

No. of Stocks	Average Turnover	Average Rebalancing Cost	Iter. $\leq 1\%$ Tracking error	Iter. $\geq 4\%$ Tracking error	OU Ratio
5	0.61	0.39%	16%	28%	1.29
10	0.58	0.54%	34%	22%	1.46
15	0.48	0.63%	19%	16%	1.29
20	0.38	0.70%	34%	16%	1.91
25	0.29	0.77%	22%	09%	2.56
30	0.25	0.85%	31%	06%	3
35	0.18	0.90%	31%	09%	2.2
40	0.11	0.94%	31%	09%	1.91
45	0.02	0.94%	19%	19%	2.56

Definitions for the columns in tables 1 and 2 : −

No. of stocks - Number of stocks in the tracker.
Average Turnover - Average of the proportion of stocks that were replaced by another set of stocks in the tracker.
Average Rebalancing Cost - Average cost in percentage terms to rebalance the tracker. The calculation includes both the proportional and fixed costs which are calculated based on the formulae in section 3. Hence, low value value of this parameter is favourable.
Iter. $\leq 1\%$ Tracking error - Proportion of tracking iteration in percentage terms which had a tracking error of 1% or less including the cases of both excess and under returns as compared to the index. Hence, high value of this parameter is favorable.
Iter. $\geq 4\%$ Tracking error - Proportion of tracking iteration in percentage terms which had a tracking error of 4% or more including the cases of both excess and under returns as compared to the index. Hence, low value of this parameter is favourable.
OU Ratio - This parameter gives the ratio of iterations in which tracker outperformed the benchmark to the iterations in which tracker underperformed the benchmark. Hence, higher OU value the better.

Results show that the cost of rebalancing goes up with increase in the number of stocks in the tracker owing to increasing fixed charges with increasing number of stocks in the tracker which have to be replaced and others which have to be re-balanced. Further, we note that tracking error does not follow a strictly increasing or decreasing trend with the number of stock. Finally, the tracking error favours the number of stocks in the 20 -30 range. We can assert this as the tracking error efficiency saturates beyond 30 stocks (also evident in figure 3) barring the case of 35 stocks in short tracker (but this tracker suffers from a poor OU ratio). Further, for less than 20 stocks, the tracking error is high which is not duly compensated with low transaction costs.
Therefore, we will be considering trackers with **20, 25 or 30** stocks in them for constructing the index.

4.2. Rebalancing frequency

Table 3 and 4 show the results obtained when we varied the rebalancing frequency for short and long trend trackers taking 20, 25 and 30 stocks in them.

Table 3: Effect of rebalancing frequency on short-trend tracker

No. of Stocks	Rebal. Frequency.	Annual Rebal. Cost	Iter. $\leq 1\%$ Tracking error	Iter. $\geq 4\%$ Tracking error	OU Ratio
20	Monthly	6%	47%	0%	1.72
	Bi-Monthly	3.24%	22%	16%	3.08
	Quarterly	2.32%	22%	28%	3
	Half-Yearly	1.36%	06%	38%	7
	Yearly	0.79%	12%	75%	7
25	Monthly	6.96%	51%	0%	2.50
	Bi-Monthly	3.78%	18%	12%	2.77
	Quarterly	2.68%	16%	28%	3
	Half-Yearly	1.52%	06%	44%	4.33
	Yearly	0.85%	12%	88%	7
30	Monthly	8.04%	56%	01%	2.06
	Bi-Monthly	4.32%	31%	12%	2.27
	Quarterly	3.00%	16%	31%	2.56
	Half-Yearly	1.66%	12%	05%	4.33
	Yearly	0.89%	12%	75%	7

Table 4: Effect of rebalancing frequency on long-trend tracker

No. of Stocks	Rebal. Frequency.	Annual Rebal. Cost	Iter. $\leq 1\%$ Tracking error	Iter. $\geq 4\%$ Tracking error	OU Ratio
20	Monthly	7.68%	59%	01%	1.13
	Bi-Monthly	4.08%	41%	06%	2.5
	Quarterly	2.80%	34%	16%	1.91
	Half-Yearly	1.48%	25%	38%	3
	Yearly	0.79%	12%	38%	7
25	Monthly	8.64%	58%	01%	1.09
	Bi-Monthly	4.56%	37%	04%	2.77
	Quarterly	3.08%	22%	09%	2.56
	Half-Yearly	1.6%	31%	44%	1.67
	Yearly	0.86%	25%	50%	7
30	Monthly	9.72%	66%	01%	1.39
	Bi-Monthly	5.1%	39%	04%	2.50
	Quarterly	3.40%	31%	06%	3
	Half-Yearly	1.94%	31%	38%	2.2
	Yearly	0.94%	12%	50%	7

We observe that although monthly rebalancing gives us extremely favorable results in terms of tracking error, it does incur exorbitant annual transaction costs which makes it undesirable and hence this frequency cannot be used. Moreover, half-yearly and yearly rebalancing are on the other end of the spectrum where they provide lower transaction costs but their tracking error is large and hence these frequencies too cannot be used.
Bi-monthly and Quarterly rebalancing acceptably balance between costs and tracking error. Quarterly rebalancing offers an edge in costs while bi-monthly rebalancing gives us better results in tracking error.

4.3. Proportion of long/short tracker in composite tracker

Table 5 shows the results obtained for different combinations of number of stocks taken to construct long-short composite tracker for 20, 25 and 30 stocks in total with bi-monthly and quarterly rebalancing window.
We observe that as the contribution of short tracker increases in the composite tracker, we get lower transaction charges but tracking error increases at the same time. The vice-versa is true for the contribution of long tracker in the composite tracker. Figure 3 depicts the relationship between transaction costs and the tracking error based on table 5. We infer that there is again no strict mono-directional relationship between the two which leads to multiple trackers being optimal for our purpose.
Figure 4 depicts the performance of four trackers which are optimal trackers marked in boldface in table 5.s

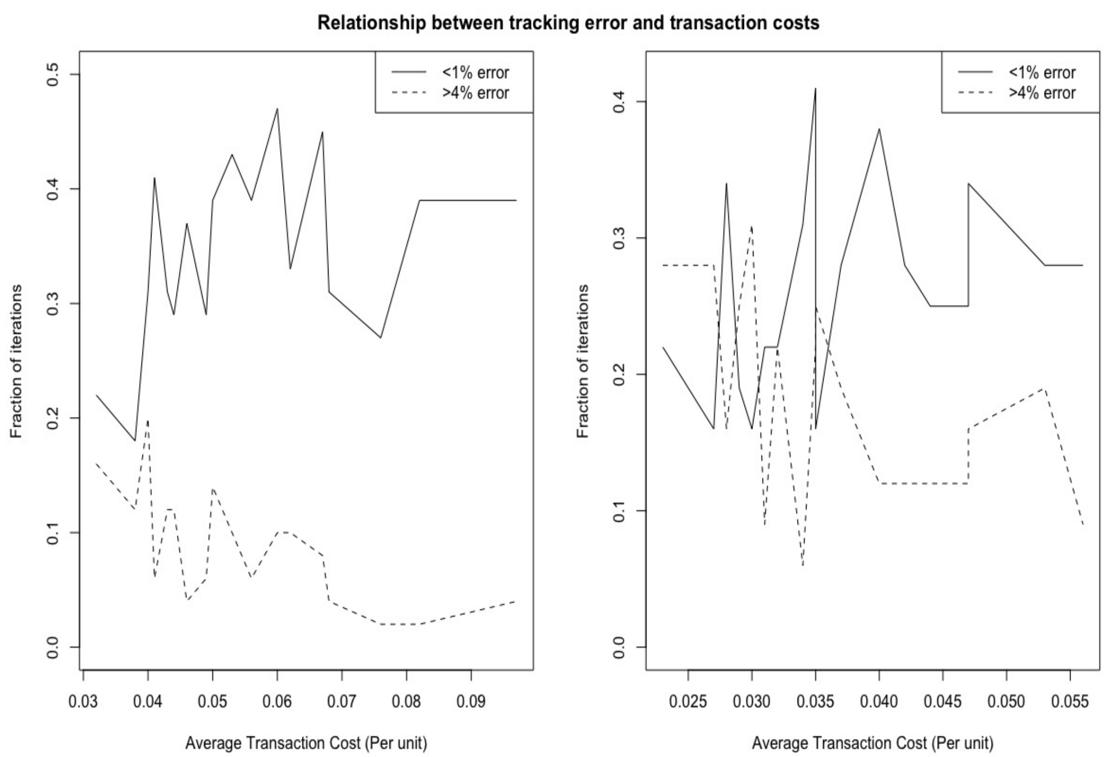

Figure 3: Left: bi-monthly rebalancing; Right: quarterly rebalancing

Table 5: Effect of tracker composition on long-short tracker with bi-monthly rebalancing

Rebal. Freq.	No. of Stocks	Composition	Annual Rebal. Cost	Iter. $\leq 1\%$ Tracking error	Iter. $\geq 4\%$ Tracking error	OU Ratio
Bi-Monthly	20	s = 0, l = 20	4.08%	41%	06%	2.5
		s = 5, l = 15	6.78%	31%	04%	1.58
		s = 10, l = 10	4.98%	39%	14%	1.45
		s = 15, l = 5	4.02%	31%	20%	1.58
		s = 20, l = 0	3.24%	22%	16%	3.08
	25	s = 0, l = 25	4.56%	37%	04%	2.77
		s = 5, l = 20	7.56%	27%	02%	2.27
		s = 10, l = 15	**6.00%**	**47%**	**10%**	**1.88**
		s = 15, l = 10	**5.28%**	**43%**	**10%**	**1.72**
		s = 20, l = 5	4.44%	29%	12%	2.06
		s = 25, l = 0	3.78%	18%	12%	2.77
	30	s = 0, l = 30	5.10%	39%	04%	2.5
		s = 5, l = 25	8.22%	39%	02%	2.5
		s = 10, l = 20	6.72%	45%	08%	3.45
		s = 15, l = 15	6.24%	33%	10%	1.88
		s = 20, l = 10	5.64%	39%	06%	2.27
		s = 25, l = 5	4.92%	29%	06%	3.08
		s = 30, l = 0	4.32%	31%	12%	2.27
Quarterly	20	s = 0, l = 20	2.80%	34%	16%	1.91
		s = 5, l = 15	4.72%	25%	12%	1.29
		s = 10, l = 10	**3.52%**	**41%**	**22%**	**1.91**
		s = 15, l = 5	2.88%	19%	25%	3
		s = 20, l = 0	2.32%	22%	28%	3
	25	s = 0, l = 25	3.08%	22%	09%	2.56
		s = 5, l = 20	5.28%	28%	19%	2.56
		s = 10, l = 15	**4.20%**	**28%**	**12%**	**2.2**
		s = 15, l = 10	3.72%	28%	19%	3.57
		s = 20, l = 5	3.20%	22%	22%	3
		s = 25, l = 0	2.68%	16%	28%	3
	30	s = 0, l = 30	3.40%	31%	06%	3
		s = 5, l = 25	5.56%	28%	09%	3.57
		s = 10, l = 20	4.72%	34%	16%	4.33
		s = 15, l = 15	4.36%	25%	12%	3
		s = 20, l = 10	4.00%	38%	12%	4.33
		s = 25, l = 5	3.48%	16%	25%	3.57
		s = 30, l = 0	3.00%	16%	31%	2.56

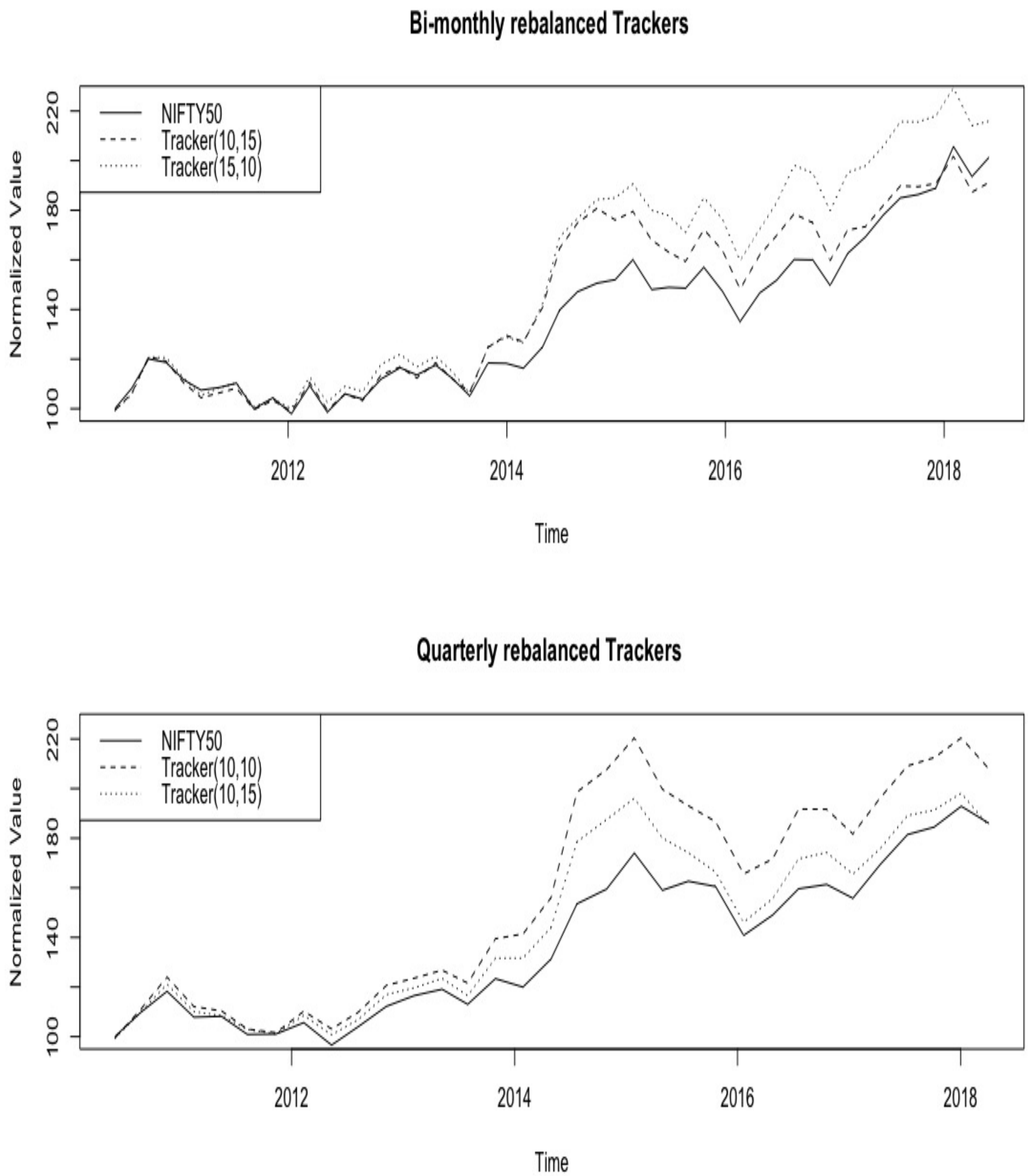

Figure 4: Performance of optimal trackers with bi-monthly and quarterly rebalancing

5. Conclusion

In this study, we proposed a method to track NIFTY50 which is one of the most representative benchmark of Indian capital markets covering over 65% of market capitalisation. We observed that we have multiple parameters at our disposal to enhance the performance of tracker. We also characterised the trade-offs involved in the process.

Further work can involve taking other benchmark indices of Indian capital markets *viz.* BSE Sensex *etc.* and implementing the aforementioned methodology on them to find the bandwidth of this procedure. Further work can also be done on dynamic weight adjustment of stocks in the tracker based on their performance in the previous iterations.

References

Nakayama, J. and Yokouchi, D. (2018). Applying time series decomposition to construct index-tracking portfolio. *Asia-Pacific Financial Markets (2018),* **25**, 341–352.

Roll, R. (1992). A mean/variance analysis of tracking error. *Journal of Portfolio Management, Summer,* **18**, 13–22.

Focardi, S. M. and Fabozzi, F. J. (2004). A methodology for index tracking based on time-series clustering. *Quantitative Finance,* **4**(**4**), 417–425.

Dose, C. and Cincotti, S. (2005). Clustering of financial time series with application to index and enhanced index tracking portfolio. *Physica A,* **355**, 145–151.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association,* **74**(**368**), 829–836.

Shibata, R. and Miura, R. (1997). Decomposition of Japanese yen interest rate data through local regression. *Financial Engineering and the Japanese Markets,* **4**, 125–146.

Thomaidis, N. S. (2013). On the application of cointegration analysis in enhanced indexing. *Applied Economic Letters,* **20**, 391–396.

Papantonis, I. (2016). Cointegration-based trading: Evidence on index tracking and market-neutral strategies. *Managerial Finance,* **42**, 449–471.

Jha, M. and Srivastava, N. (2013). Portfolio rebalancing model using fuzzy optimization. *International Journal of Scientific Engineering and Research (IJSER),* **1**(**4**), 2347-3878.

Kellerer, H., Mansini, R. and Speranza, M. G. (2000). Selecting Portfolios with fixed costs and minimum transaction lots. *Annals of Operations Research,* **99**, 287–304.

NIFTY 50 Index Methodology. Retrieved on 20 March 2020 from www1.nseindia.com/content/indices/Method_Nifty_50.pdf.

APPENDIX

1. Calculation of proportional component of transaction cost

In calculation of the transaction charges per unit of capital. Assume the following :-
n - number of stocks in tracker.
l - number of stocks to be dropped/added after rebalancing window
$\frac{1}{n}$ - Initial investment in each stock.
X_i - Final value of the investment in stock i.
R_i - Return on stock i $(\frac{X_i}{\frac{1}{n}})$.
$\sum_{i=1}^{n} X_i$ - Total value of investment at the end of rebalancing window.
p - proportional rebalancing cost per unit MAD - Mean Absolute Deviation

Now, the amount that should be present in each stock after the rebalancing period is :-
$\frac{\sum_{i=1}^{n} X_i}{n} = \bar{X}$
Hence, transaction cost will be given as:-
$Cost = p\sum_{i=1}^{n} |X_i - \bar{X}| = pnMAD(X_i)$

But we will have to reduce from this value the contribution of those stocks which will be replaced after the transaction period. Therefore,

$Cost = pnMAD(X_i) - \sum_j |X_j - \bar{X}|$, where j denotes the set of stocks to be dropped

Further, we add the transaction cost of dropping those stocks and adding new stocks to the portfolio.

Cost of dropping $= p\sum_j X_j$

Cost of adding $= p\sum_1^l \bar{X}$

Hence, total proportional cost is given by :-

$Cost = pnMAD(X_i) - \sum_j |X_j - \bar{X}| + p\sum_j (X_j + \bar{X})$
$= pnMAD(X_i) + 2p\sum_j min(X_j, \bar{X})$ by identity $(a + b - |a - b| = 2min(a, b))$

$= pMAD(R_i) + 2p(\frac{l}{n})(\frac{1}{l}\sum_j min(R_j, \bar{R}))$

$= pMAD(R_i) + 2p(turnover)(\frac{1}{l}\sum_j min(R_j, \bar{R}))$