February 27, 2015

Technical S&P500 Factor Model

A single unified technical factor based model that has consistently outperformed the S&P Index

By Manish Jalan

The paper describes the objective, the methodology, the back-testing and finally the results of building a single unified technical factor model which has consistently worked in the US stock market over the last 10 years. The factor model has been built by rigorous testing and analysis of technical factors spanning across the S&P 500 stocks trading in the US Stock Exchange.

1. The Objective

The objective of building the S&P 500 unified factor model was to identify key technical factors which have been working consistently in the US market irrespective of the bull or bear market cycles. The second objective was to identify what weightages each of the individual factor shall carry so that the overall portfolio consistently generates superior alpha. The third objective was to come up with a single unified factor which was a combination of several factor and can be used to rank the stocks from 1 to 500. The objective of ranking the stocks was to ensure that the top decile stocks would consistently outperform the broader market and the bottom decile stocks would consistently underperform the broader market. Hence, the final unified factor would be of the form:

Final Unified Factor = A*Factor1 + B*Factor2 …

Where, A, B are the weightages to each factor such that A+B+… = 1
Factor1, Factor2… etc. are technical factors like 3 month stock return, Momentum of the stock etc.

The overall objective was to find a single unified factor which has been stable and generated consistent alpha in the US market over the last 11 years.

2. The Mechanism

The mechanism for identifying the key factors was based on monthly re-balance of stocks. Historical back-test was carried from beginning of Jan 2004 to Oct 2014 period. The stocks were ranked at the 1st trading day of each month from 1 to 500. The ranking was based upon factor weightage which means that higher is the value of a factor for a given stock higher is its overall rank in the portfolio of 500 stocks.

The returns of the top ranked decile stocks (based on highest factor weight for the month) were compared to the returns of the bottom decile stocks for the same month. Optimization and Monte-Carlo simulations was then carried out to identify that for what combination of factors, shall the top decile stocks consistently outperform the bottom decile stocks on a month on month basis with the outperformance Sharpe ratio greater then 1 (or highest achievable Sharpe ratio) for the period Jan 2004 to Oct 2014.

3. The Data & Assumptions

The data for the historical back-test and optimization comprised of daily closing prices of all 500 stocks from 2000 onwards. The data was thoroughly cleaned and adjusted for stocks splits, bonuses and dividends.

The assumptions made while building the model were that each stock in the 500 stocks universe are fairly liquid and there is no additional liquidity filter required to filter the stocks. The factors shall work uniformly across the breadth of the stock, and will not be biased based on the stocks liquidity or turnover volumes. Secondly the factor model has been built with specific objective of identifying top decile stocks which can consistently outperform the S&P Index on a month on month basis.

4. The Historical Back-test

The back-testing of the factors comprised of following critical steps: the data sampling, the factor identification, z-scoring of the factors, identifying alpha generating factors and running historical trade analytics.

4.1 The Factor Identification

The first step to build the model was to identify which factors, should be included in the overall analysis. As a first step it was imperative that we wanted to include a combination of momentum and mean reversion factors – so that the overall model is not biased towards any one given factor or any one given market circumstances. E.g.: In 2002-2006 periods the momentum based companies had done very well but in 2008 the momentum based stocks grossly underperformed the S&P Index. Similarly momentum factors might do very well in trending markets like 2002 to 2008 but in range bound markets like 2010-2011, the mean reversion factors might outperform the broader market.

Keeping these issues in mind the following factors were shortlisted for testing:
Technical factors: A total of 28 factors were identified on the technical side, based on the price and volume action of the stocks. The idea was to shortlist as many non-correlated, high performing factors as possible, which could then be used to construct the single unified factor. The Table 1 describes the factor name, the factor type and factor description / formula which were used for the technical side of factor testing.

Factor Name	Factor Type	Factor Description
SlopeWeekly	Momentum	Price slope of 10 Week Exponential Moving Average (EMA) over 5 weeks
Volumentum-weekly	Momentum	[Price (End of this week) - Price (End of last week)]*[Avg Week Volume / Avg 6 Mo Volume]
Volumentum-monthly	Momentum	[Price (End of this month) - Price (End of last month)]*[Avg Monthly Volume / Avg 12 Mo Volume]
Momentum-3Mo	Momentum	Avg of daily returns of last 3 months
Momentum-6Mo	Momentum	Avg of daily returns of last 6 months
Momentum-9Mo	Momentum	Avg of daily returns of last 9 months
Mean-reversion -5-250	Mean Reversion	(Price Avg for 5 Days - Price Avg for 250 Days)/Price Avg for 250 Days
Mean-reversion -5-500	Mean Reversion	(Price Avg for 5 Days - Price Avg for 500 Days)/Price Avg for 500 Days
Mean-reversion -5-1000	Mean Reversion	(Price Avg for 5 Days - Price Avg for 1000 Days)/Price Avg for 1000 Days
HighLowRange	Momentum	(Current price - 52 week price low)/(52 Week High – 52 Week Low)
MoneyFlow	Momentum	Money Flow = (((Close-Low) - (High-Close)) / (High-Low)) * Volume
MoneyFlowPersistency 1 Month	Momentum	No of days when Money Flow was positive in 1 months / Number of Days in 1 months
MoneyFlowPersistency 3 Month	Momentum	No of days when Money Flow was positive in 3 months / Number of Days in 3 months
MoneyFlowPersistency 6 Month	Momentum	No of days when Money Flow was positive in 6 months / Number of Days in 6 months
SlopeDaily	Momentum	Price slope of 10 Day Exponential Moving Average (EMA) over 5 days
SlopeMonthly	Momentum	Price slope of 10 Month Exponential Moving Average (EMA) over 5 months
3YrRet	Momentum	Price return in percentage in 3 years
30DayRet	Momentum	Price return in percentage 30 days
60DayRet	Momentum	Price return in percentage 60 days
90DayRet	Momentum	Price return in percentage 90 days
3YrCurrPxRet	Momentum	(Current Price – Moving Avg of Last 3 yrs Price) /Current Price
30DayADP	Momentum	Avg of daily returns of last 3 months
60DayADP	Momentum	Avg of daily price change of last 60 days
90DayADP	Momentum	Avg of daily price change of last 90 days
-0.5*3YrRet+0.5*30DayRet	Mean Reversion	-50% of 3 year price return + 50% of 30 days price return
-0.5*3YrRet+0.5*60DayRet	Mean Reversion	-50% of 3 year price return + 50% of 60 days price return
-0.5*3YrRet+0.5*90DayRet	Mean Reversion	-50% of 3 year price return + 50% of 90 days price return
-0.5*3YrAvgCurrPxRet+0.5*SlopeDaily	Mean Reversion	-50% of 3YrCurrPxRet + 50% of SlopeWeekly

Table 1: Technical factors that were used in building the model

After the identification of each factor, the factors were calculated using Java programs for each stock at the end of each month (from 2004 to 2014 on a month on month basis). The idea was that if the stocks need to be re-balanced on 1st trading day of each month, then the factors should be calculated at the end of each previous month to derive the ranking of the stocks, which can be re-balanced the next trading day.

4.2 The lateral Z-Scoring of Factors

Combining of the factors is usually a key challenge in factor modeling. For E.g.: simply saying, 0.5*HighLowRange + 0.5*SlopeWeekly makes no sense at all because HighLowRange and SlopeWeekly are different set of data which cannot be linearly combined. Hence, as an important exercise after calculation of the factors, for each stocks on a month on month basis a lateral z-scoring of each factor was done across the entire universe of 500 stocks (or the active stocks for that particular month) so that the factors can be converted into their respective z-scores.

As an example assuming that on 30th April'2002 the HighLowRange of 500 active stocks were HLR1, HLR2….HLR500. Then the lateral z-score of Stock 1 on the day would be calculated as:

HLR1_ZSCORE = (HLR1 – (Average (HLR1, HLR2…HLR789)) / Stdev (HLR1, HLR2…HLR789)

Of stocks2 will be:

HLR2_ZSCORE = (HLR2 – (Average (HLR1, HLR2…HLR789)) / Stdev (HLR1, HLR2…HLR789)

And so on.

4.3 The Data Sampling

The third important step was to sample the entire data set from 2004 to 2014 into in-sample and out-of-sample periods. The objective was to avoid data fitting and development of a model which can work efficiently even on a blind set of data. For the first set of back-test only the first 70% of the data (in-sample data) was used. Hence, historical data from 2004 to 2010 was used for identifying the alpha generating factors. Once the factors which have worked efficiently in the period 2004-2010 were identified, these factors were extended to blind set of data from 2011 to 2014 periods (out of sample data).

Although many possible variants of in-sampling and out-of-sampling was available like testing for alternative years, testing for first five years and applying to next 3 years and so on, but due to availability of large data pool spanning 11 years, the 70-30 ratio of in-sample to our-of-sample managed to capture most of the market dynamics which the model could likely encounter in the future.

4.4 Identifying alpha generating factors

Once the factors were identified, the lateral z-scoring and data sampling was done, the next step was to generate the trades. As per the assumption of re-balancing at the beginning of each month, all the trades were generated from 2004 to 2014 (Separately for in-sample and out of sample data). Hence, a typical trade would comprise of a stock name, stock buying date (beginning of the current month), stock selling date (beginning of next month), trade return generated by holding the stocks for 1 month, Factor1_zscore,Ffactor2_zscore … FactorN_zscore. Note that the factors z-score were taken as on 1 day prior to the re-balance day as we wanted to identify the power of the factor in predicting the future 1 month performance of the stock.

Overall a total of 66,000 trades were generated by our Java programs for 12 months * 11 Years * 500 (average active stocks).

The trade returns were then regressed against the factors: Factor1_zscore, Factor2_zscore … FactorN_zscore on a year on year basis using R statistical package. Hence we would take the returns for 1 year, say 2004 and do a regression of the trade returns against all the factors in focus (28 technical factors). This would yield a single t-stat for each factor for a given year. The absolute value of t-stat would indicate how important that factor has been in that year to identify the next 1 month returns of the stock. A positive value of t-stat means that higher that factor is higher is the trade return and a negative t-stat signifies that lower the factor is higher is the trade return. Table 3, shows the average t-stats, stdev in the t-stats and the Sharpe ratio of t-stats (average / stdev) of all the factors for the period of 2004 to 2014 (as analyzed for all 500 stocks in the universe).

Overall Analysis	Overall t-stats
SlopeWeekly	2.46
Volumentum-weekly	1.09
Volumentum-monthly	1.08
Momentum-3Mo	1.21
Momentum-6Mo	2.68
Momentum-9Mo	2.16
Mean-reversion -5-250	2.68
Mean-reversion -5-500	1.66
Mean-reversion -5-1000	1.80
HighLowRange	-1.63
MoneyFlow	-0.83
MoneyFlowPersistency 1 Month	2.52
MoneyFlowPersistency 3 Month	0.23
MoneyFlowPersistency 6 Month	-2.01
SlopeDaily	2.53
SlopeMonthly	-0.22
3YrRet	1.84
30DayRet	2.36
60DayRet	3.59
90DayRet	3.74
3YrCurrPxRet	2.64
30DayADP	2.24
60DayADP	3.21
90DayADP	2.73
-0.5*3YrRet+0.5*30DayRet	-1.67
-0.5*3YrRet+0.5*60DayRet	-1.29
-0.5*3YrRet+0.5*90DayRet	-0.91
-0.5*3YrAvgCurrPxRet+0.5*SlopeDaily	3.64

Table 2: The t-stats and consistency in t-stats of the factors tested in US market from 2004 to 2014

Please note that although the in-sample (2004-2010) and out-of-sample (2011-2014) factor analysis was done separately, we present below the t-stats achieved across the entire set of data. It was coherently found that factors which worked well in in-sample period of 2004-2010 (higher t-stats factor) also continued to work consistently in out-of-sample period of 2011-2014 and hence avoiding the need to work further on factors which could work first in in-sample period and later extend to out-of-sample period. It is beyond the scope of paper to present a detailed study and breakup of in-sample and out-of-sample studies.

Referring to Table 3, it is quite evident that the most consistent Sharpe in t-stats has been of the factors SlopeWeekly (2.46) , Momentum-6Mo (2.68), Mean-reversion -5-250 (2.68), HighLowRange (-1.63) , MoneyFlowPersistency 1 Months (2.52), MoneyFlowPersistency 6 Months (-2.01), Day90Ret (3.74), -0.5*3YrAvgCurrPxRet+0.5*SlopeDaily (3.64) and so on.

There were many other factors which had t-stats of more than 1 and were quite significant. However, after running the correlation study (using correlation matrix) – some of these factors were very highly correlated to each other. The overall objective of the exercise was to find as many "non-correlated" factors as possible – which had the highest t-stats possible. We also carried out several step-wise regressions to reduce the number of factors and come up with the most significant ones.

Hence after a series of correlation study and step-wise regression the factors which were found to be of most significance were:

Short Term Technical Factors: SlopeWeekly, MoneyFlowPersistency1Months, Day90Ret

Long Term Technical Factors: HighLowRange, MoneyFlowPersistency6Months, MR3YrPriceSlopeDaily

Please note that the other significant factors like Momentum-6Mo, Mean-reversion -5-250 etc. although had higher t-stats they all either had high correlation with our top technical factors or in the later stages of monte-carlo failed to yield superior results when combined with other factors. Hence, the final 6 factors (3 from short term technicals and 3 from medium/long term technicals) have been identified through a rigorous set of optimization and t-stats analysis, describing all of which is beyond the scope of this paper.

5. Monte Carlo Simulation

Having identified the top 6 factors, in the last section the next step was to combine the 6 factors and give them an appropriate weightage to come up with a unified factor. Monte-carlo techniques was developed using Java programs, whereby the weight of each of the factor was varied from 0% to 50% in steps of 25% each. The objective was to get a set of weights for each of the 6 factors such that the top decile stocks could consistently outperform the S&P Index. Hence, for a given month stocks would be ranked based on the unified factor value. Higher the factor value – higher would be the rank of the stock. The average return of the portfolio for the month would then be calculated as:

Average Portfolio Return for month = Average return of top decile ranked stocks – Average return of the S&P Index

The objective of the Monte-carlo simulation is to then maximize this average portfolio return and the Sharpe ratio of this return on an annualized basis. Hence the final unified factor which got derived was:

Final Factor: 0.25* SlopeWeekly-0.5* HighLowRange+0.5* MoneyFlowPersistency1Months-0.5* MoneyFlowPersistency6Months+0.5* Day90Ret+0.5* MR3YrPriceSlopeDaily

Final Unified Factor: **+0.25* SlopeWeekly**
-0.5* HighLowRange
+0.5* MoneyFlowPersistency1Months
-0.5* MoneyFlowPersistency6Months
+0.5* Day90Ret
+0.5* MR3YrPriceSlopeDaily

The above single unified factor has worked by far the most consistently in the S&P 500 stocks over the last 11 years. A positive sign of mean reversion, momentum and growth weights signifies that higher are these values higher is the expected forward 1 month return of the stocks. A negative weight of value factor signifies lower this value is higher is the forward 1 month return of the stock.

Rationale behind the technical factor model is also quite unique. The table below shows the kind of stocks which gets picked up by the technical factor model.

Overall Nature of the stocks pick-up by the model
Long Term: Negative Money Flow, Near to the bottom to middle of its high low range, Had a negative returns in long term (> 252 Days)

Short Term: Positive weekly slopes, Positive money flow, Good daily returns - short term positive
Hence the factor model tries to capture stocks which in the long term has been left out by the investors and in the short term the momentum has just started building in the stocks

6. The Results

The rationale of picking stocks which in the long term has been underperforming the S&P Index and in the short term has started building positive momentum means that we are picking undervalued stocks which tends to work in most market environments.

The average monthly return of the top decile stocks from 2004 to 2014 stood at 1.83% and the average monthly return of the S&P Index was a mere 0.51%. Hence on an average the top decile stocks outperformed the S&P Index by a whopping 1.32% on a month on month basis.

Hence, if a portfolio manager had consistently bought the top decile stocks in the technical factor basket then he would have been able to outperform the S&P Index by 15.8% annually. Please note that to keep things simple and in perspective we have not considered the transaction costs and brokerage costs in buying and selling of the stocks.

Fig 1, shows the YoY performance of the top decile stocks as compared to the S&P Index. It is quite notable that even in the out of sample period of 2011 to 2014 – the outperformance to the index occurred in 2011, 2012 and 2013 - 3 out of the 4 years. Hence, the results are not biased towards in-sample period and works consistently across different market environments.

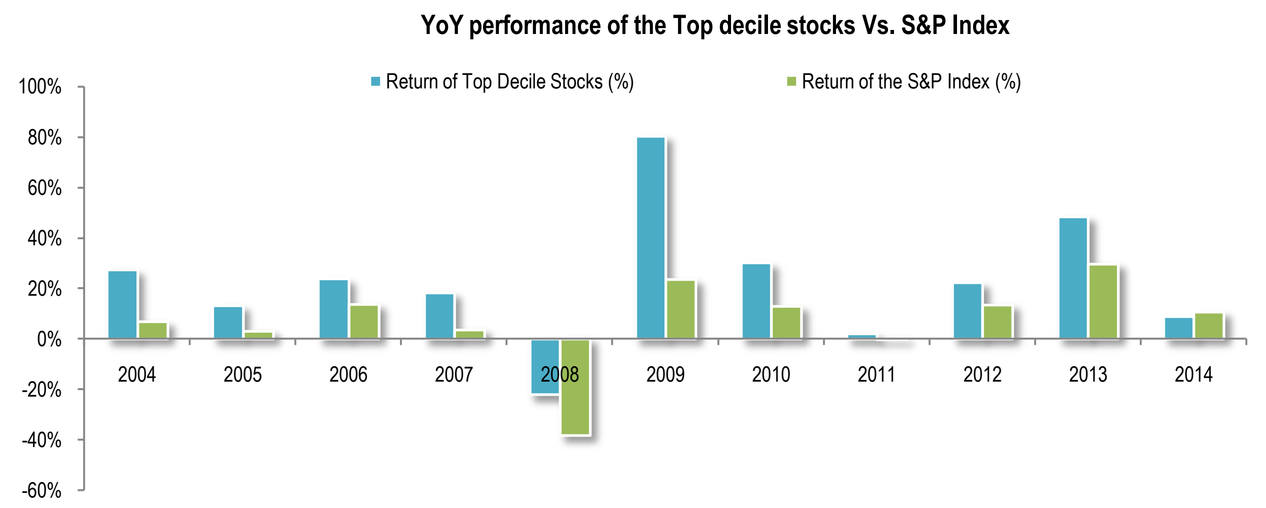

Fig 1: Year On Year Performance comparison of the Top Decile Stocks in Factor Basket Vs. S&P Index returns

If a fund manager were to construct a portfolio based on the technical factor model , where he was long the top decile stock and he would re-balance the portfolio on a monthly basis then the performance of the fund for the last 11 years would look like that in Table 3.

Year	Annualized Return (Factor Basket)	Annualized Return (S&P Index)
2004	27.16%	6.68%
2005	13.13%	3.00%
2006	23.70%	13.62%
2007	18.24%	3.53%
2008	-22.21%	-38.49%

2009	80.29%	23.45%
2010	29.97%	12.78%
2011	1.97%	0.00%
2012	22.14%	13.41%
2013	48.27%	29.60%
2014	8.78%	10.44%
Avg Annual Return	**22.86%**	**7.09%**
Vol in Annual Return	26.07%	17.49%
Sharpe Ratio	**0.88**	**0.41**

Table 3: Annualized returns of the technical factor basket and the S&P Index. Shows that technical factor basket outperformed the S&P Index by 15.77% annually with a Sharpe ratio of 0.88

The cumulative return of NAV of 100 USD at the beginning of Year 2004 would stand at 770.47 by Oct'2014! The graph of the cumulative NAV (assuming no transaction cost, slippages, cost of capital etc.) is as shown in Fig 2.

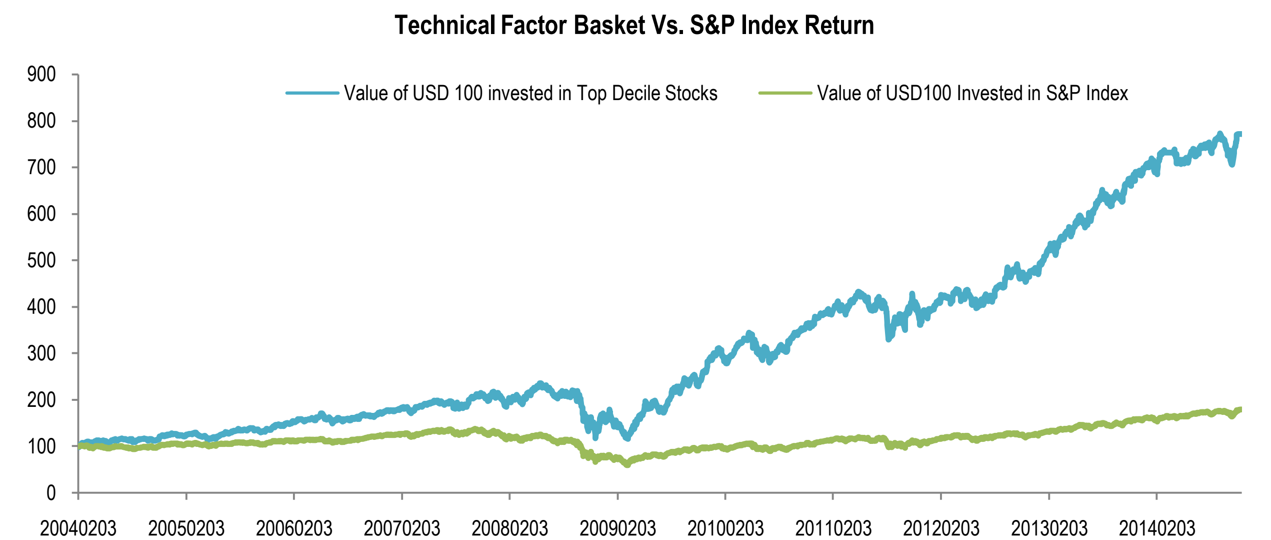

Fig2: Cumulative value of USD 100 invested in Jan 1999 in a Long/Short portfolio constructed using the unified factor model in US market and re-balanced on monthly basis (assuming no transaction costs)

As evident from Fig 2, the portfolio constructed by using the technical factor and choosing the top decile stocks would outperform the S&P Index in 10 out of 11 years. Also the portfolio would have cumulatively yielded a return of 770 (for USD 100 invested in 2004) as opposed to S&P Index yielding and return of USD 179 (for USD 100 invested in 2004).

7. Conclusion

The objective of this study was to narrow down the factors which has best worked in US market from a large universe of 30+ factors to a smaller set of under 5 factors.

Firstly the 28 factors were chosen from a diversified set of mean reversion and momentum. Data was first broken into in-sample (2004-2010) and out-of-sample (2011-2014) periods. A historical back-test for the last 11 years from 2004 to 2014 was carried out to identify the most alpha generating factors. The factors identified after a rigorous back-testing were such that they have consistently performed well in the US market (both in in-sample and out-of-sample periods) over the last 11 years and is not biased towards a given kind of market environment like bull, bear or range bound market. The top 6 factors which were finally shortlisted were:

Short Term Technical Factors: SlopeWeekly, MoneyFlowPersistency1Months, Day90Ret

Long Term Technical Factors: HighLowRange, MoneyFlowPersistency6Months, MR3YrPriceSlopeDaily

The monte-carlo simulation was then carried out to come up with a single unified technical factor. The final result of monte-carlo simulation showed that the final technical factor model that gives the best set of result belongs to choosing technical factors such that:

Stocks with long term underperformance and short term to medium term outperformance gives the best set of results

The overall results of the factor model shows that the top decile stocks outperform the S&P Index by an average of about 1.32% per month and 15.77% annualized. The annualized Sharpe ratio of Top Decile stocks are 0.88 as compared to S&P Index having a sharpe ratio of 0.41. A fund constructed out of being long the top decile stock in equal dollar value would have yielded an average annualized return of a whopping 22.86% (without any transaction cost, slippages and cost of capital) and an annualized Sharpe of 0.81 with only 1 year – 2008 being negative.

Overall it can be concluded that using techniques of sound historical back-testing, monte-carlo simulation and filtering from a universe of 28+ factors to 6 factors has yielded significantly improved and diversified factors which has stayed stable and generated consistent alpha in the US market over the last 11 years.

8. Extension of factor modeling to other areas (ETFs, Asset Allocation etc)

The concepts of factor modeling on technical and fundamental parameters can be applied to various other products like ETF baskets and asset allocations. Based on numerous sets of technical and fundamental factors a basket of ETFs can be selected which can outperform a given benchmark. Similarly dynamic asset allocations between say Gold, Equities and Fixed Income can be constructed such that based on technical factors we can identify which asset to be overweight and which asset to be under weight. Rigorous back-tests can be carried out on the historical data to identify the best set of ETF basked to hold or assets to be over weight in – so that a targeted risk reward portfolio is made available to the investors.

About Us

Samssara Capital Technologies LLP ("Samssara") is a niche quant analytics firm providing end-to-end services in the areas of quantitative trading, investment, optimization and analytics space to clients globally.

The team at Samssara works on mathematical models and statistical tools that identify repetitive patterns in equity, commodity currency and treasury markets globally. We offer solutions to ride the volatility of the markets and generate consistent returns with systematic approach.

Samssara was founded in 2010 by a highly professional and experienced team of alumni of IIT Bombay.

Contact Us

Samssara Capital Technologies LLP
602, Vakratunda Corporate Park
Vishveshwar Nagar
Goregaon (E), Mumbai 400 063, India

D: +91 22 6671 9198
E: manish@samssara.com, tarun@samssara.com
W: www.samssara.com