Long Short Factor Model in HK Market

A single unified long short factor model that has worked consistently in Hong Kong stock market

By Manish Jalan

March 10, 2015

The paper describes the objective, the methodology, the back-testing and finally the results of building a single unified factor model which has consistently worked in the HK stock market over the last 13 years. The factor model has been built by rigorous testing and analysis of technical and fundamental factors spanning across the 1000 most liquid stocks trading in the HK stock exchange.

1. The Objective

The objective of building the HK 1000 unified factor model was to identify key technical and fundamental factors which have been working consistently in the HK market irrespective of the bull or bear market cycles. The second objective was to identify what weight ages 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 1000. 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 and fundamental factors like 3 month stock return, PE ratios etc.

The overall objective was to find a single unified factor which has been stable and generated consistent alpha in the HK market over the last 13 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 1999 to Aug 2011 period. The stocks were ranked at the 1st trading day of each month from 1 to 1000. 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 1000 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 1999 to Aug 2011.

3. The Data & Assumptions

The data for the historical back-test and optimization comprised of daily closing prices of all 1000 stocks from 1999 onwards. The data was thoroughly cleaned and adjusted for stocks splits, bonuses and dividends. The fundamental data for the stocks comprised of the quarterly balance sheet and income statement data which was available for most companies from the year 2000 onwards.

The assumptions made while building the model were that each stock in the 1000 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 and bottom decile stocks on a month on month basis and does not hold any analysis on outperformance w.r.t to Hang Seng Indices. Although a separate analysis on outperformance of top decile stocks vis-à-vis Hang Send indices can be carried out – but that is beyond the scope of current study.

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, mean reversion, growth and valuation 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 growth based companies had done very well but in 2008 the growth based companies underperformed the high value stocks. 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 29 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	Momentum	No of days when Money Flow was positive in 6 months / Number of Days in 6 months
RelativePrice	Momentum	6 month price performance relative to universe
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
ObvSlopeWeekly	Momentum	5 week slope of the OBV line (OBV = On Balance Volume)
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
WeeklySlopeVol	Momentum	(5 Week Slope) / Volatility in Weekly Price
-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

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

Fundamental factors: The fundamental factors identified were in the value and growth categories. Since many of the fundamental factors tend to have a high correlation (E.g.: High net income of a company usually leads to higher ROE on a year on year basis), hence the fundamental factors were selected such that they were intuitive as well as can take care of most of the company performance figures accurately. Table 2 highlights the key non-correlated fundamental factors which were used in the analysis.

Factor Name	Factor Type	Factor Description
EV/EBITDA	Valuation	EV = ((close * all issue cap shares) + liability total + minority interest + share cap pref – cash) EBIDTA score = (PBT – associates + CF interest received – fin. costs + depreciated amort.)
ROE	Growth	Return on Equity
PE	Valuation	PE = Price to Earnings Ratio (Price / EPS)

Table 2: Fundamental factors that were calculated 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 1999 to 2001 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*ROE makes no sense at all because HighLowRange and ROE are different set of data which cannot be linearly combined. Hence, as an important exercise after calculation of the factors what that, for each stocks on a month on month basis a lateral z-scoring of each factor was done across the entire universe of 1000 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 789 active stocks were HLR1, HLR2….HLR789. 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 1999 to 2011 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 1999 to 2008 was used for identifying the alpha generating factors. Once the factors which have worked efficiently in the period 1999-2008 were identified, these factors were extended to blind set of data from 2009 to 2011 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 13 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 1999 to 2011 (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 124,800 trades were generated by our Java programs for 12 months * 13 Years * 800 (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 1999 and do a regression of the trade returns against all the factors in focus (29 technical and 3 fundamental 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 1999 to 2011 (as analyzed for all 1000 stocks in the universe).

Please note that although the in-sample (1999-2008) and out-of-sample (2009-2011) 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 1999-2008 (higher t-stats factor) also continued to work consistently in out-of-sample period of 2009-2011 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 -0.5*3YrRet_zscore + 0.5*90DayRet_zscore (1.74) followed by 30DayRet_zscore (1.22) , 60DayRet_zscore (1.17), 90DayRet_zscore (1.17), SlopeMonthly_zscore (1.16), HighLowRange_zscore (0.96) and so on.

Since taking the top 10 factors (or all factors with Sharpe of t-stats > 1) in the final unified factor only complicates the analysis, we ran a correlation study of top 10 factors against each other to reduce the dimensionality of factors.

It was identified that the top factor -0.5*3YrRet_zscore + 0.5*90DayRet_zscore had a correlation of 0.32, 0.48 and 0.62 with 60DayRet_zscore, 90DayRet_zscore respectively. It has a mere 0.0027 correlation with SlopeMonthly_zscore.

Hence the top 2 factors in technical factors with near zero correlation was

Mean reversion factor: -0.5*3YrRet_zscore+0.5*90DayRet

Momentum factor: SlopeMonthly_zscore

Similarly it was identified that the top 2 factors in fundamental factor with near zero correlation was:

Value factor: EV/EBITDA_zscore

Growth factor ROE_zscore

with a correlation of 0.0023 between them. Please note that the t-stat of EV/EBITDA_zscore factor is negative (-0.89) which is quite intuitive, as in the long term stocks with lower Enterprise Value per EBIDTA income tend to outperform the broader index.

Please note that the other significant factors like HighLowRange_zscore, Volumentum-weekly_zscore etc. although had higher Sharpe of t-stats they all either had high correlation with our top 2 technical factors or in the later stages of monte-carlo failed to yield superior results when combined with other factors. Hence, the final 4 factors (2 each in technical and fundamental sides) have been identified through a rigorous set of optimization and t-stats analysis, describing all of which is beyond the scope of this paper.

Overall Analysis	Average t-stats (per year)	Stdev in t-stats (per year)	Sharpe / Consistency in t-stats
SlopeWeekly_zscore	0.86	1.22	0.71
Volumentum-weekly_zscore	1.24	3.45	0.36
Volumentum-monthly_zscore	-0.41	2.46	-0.17
Momentum-3Mo_zscore	1.41	6.01	0.23
Momentum-6Mo_zscore	1.41	6.54	0.22
Momentum-9Mo_zscore	0.89	6.44	0.14
Mean-reversion -5-250_zscore	0.24	3.26	0.07
Mean-reversion -5-500_zscore	1.06	3.21	0.33
Mean-reversion -5-1000_zscore	0.50	2.51	0.20
HighLowRange_zscore	2.27	2.35	0.97
MoneyFlow_zscore	-0.52	1.72	-0.30
MoneyFlowPersistency_zscore	0.98	2.31	0.42
RelativePrice_zscore	0.32	3.04	0.10
SlopeDaily_zscore	0.51	1.29	0.40
SlopeMonthly_zscore	1.65	1.41	1.17
3YrRet_zscore	-0.99	1.65	-0.60
30DayRet_zscore	1.95	1.59	1.22
60DayRet_zscore	1.65	1.41	1.17
90DayRet_zscore	1.49	1.27	1.18
3YrCurrPxRet_zscore	0.68	2.70	0.25
ObvSlopeWeekly_zscore	-0.47	1.26	-0.38
30DayADP_zscore	1.96	5.60	0.35
60DayADP_zscore	1.01	6.28	0.16

90DayADP_zscore	1.11	6.46	0.17
WeeklySlopeVol_zscore	1.11	2.56	0.43
-0.5*3YrRet_zscore+0.5*30DayRet_zscore	1.42	1.48	0.96
-0.5*3YrRet_zscore+0.5*60DayRet_zscore	1.31	1.39	0.94
-0.5*3YrRet_zscore+0.5*90DayRet_zscore	1.85	1.06	1.74
-0.5*3YrAvgCurrPxRet_zscore+0.5*SlopeWeekly_zscore	0.19	1.67	0.11
EV/EBITDA_zscore	-0.55	0.62	-0.89
ROE_zscore	1.05	1.52	0.69
PE_zscore	-0.29	0.88	-0.32

Table 3: The t-stats and consistency in t-stats of the factors tested in HK market from 1999 to 2011

5. Monte Carlo Simulation

Having identified the top 4 factors, in the last section the next step was to combine the 4 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 5% to 50% in steps of 5% each. The objective was to get a set of weights for each of the 4 factors such that the overall Sharpe ratio of the difference between top and bottom decile stocks on a month on month basis could be maximized. 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 bottom decile ranked stocks

The objective of the Monte-carlo simulation to then to maximize this average portfolio return and the Sharpe ratio of this return on an annualized basis. The results of Monte-carlo simulation were quite interesting as the Sharpe of the portfolio returns was maximized when all the 4 factors got equal weightage. Hence the final unified factor which got derived was:

Final Factor: 0.25* Mean Reversion Factor + 0.25* Momentum Factor - 0.25*Value Factor + 0.25*Growth Factor

Final Unified Factor: 0.25*[-0.5*3YrRet_zscore+
0.5*90DayRet_zscore]
+ 0.25*SlopeMonthly_zscore
- 0.25*EV/EBIDTA_zscore
+ 0.25*ROE_zscore

The above single unified factor has worked by far the most consistently in the HK 1000 stocks over the last 13 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 that lower is the EV/EBIDTA of the company higher is its forward 1 month return!
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.

6. The Results

The single unified factor with equal weightage to each diversified factor basically means that the combination of these factors works overall in most of the market environment and is not biased towards momentum, mean reversion or value/growth market environments.

The average monthly return of the top decile stocks from 1999 to 2011 stood at 2.94% and the average monthly return of the bottom decile stocks were a mere 0.46%. Hence on an average the top decile stocks outperformed the bottom decile stocks by a whopping 2.48% month on month.

The volatility in this outperformance was 7.63% on a month on month basis. The Sharpe of outperformance is 1.13 on an annualized basis. The top decile outperformed the bottom deciles in 68.42% of the months. The bar graphs on difference between top and bottom decile stocks shows the consistency with which the factor has been working over the last 13 years.

If a fund manager were to construct a portfolio based on the single unified final factor, where he was long the top decile stock and short the bottom decile stock (a dollar neutral Long/Short portfolio) and he would re-balance the portfolio on a monthly basis then the performance of the fund for the last 10 years would look like that in Table 4.

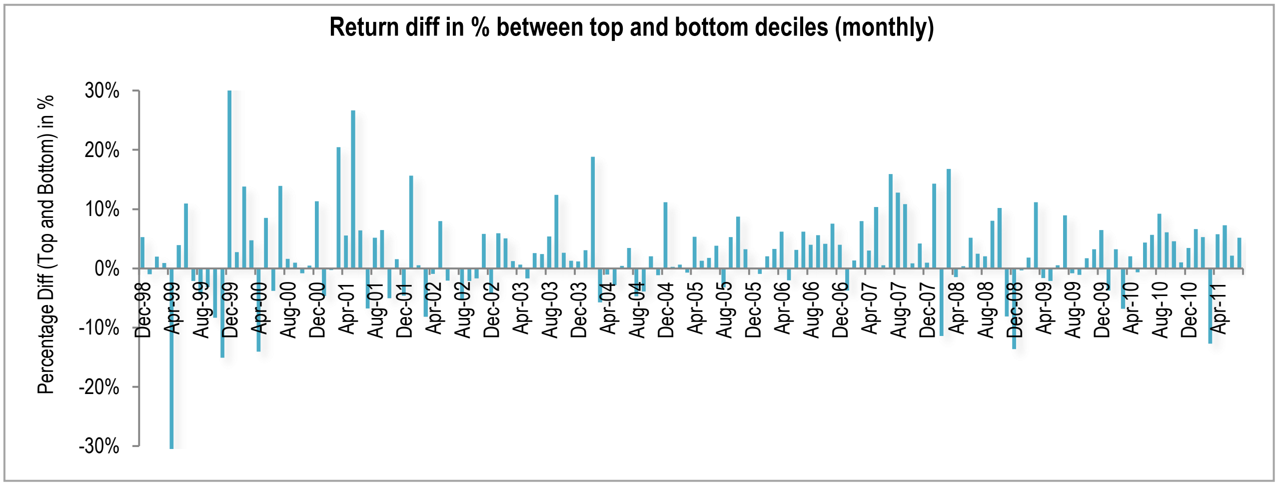

Fig 1: The average difference in returns in percentage between top and bottom decile stocks – rebalanced on monthly basis

Annualized Numbers	Values
Average Annualized Return (Assuming no slippages, no transaction costs, no cost of capital etc.)	36.10%
Annualized Stdev in Return	21.66%
Annualized Sharpe	1.67
Percentage of Positive Years	100.00%
Worst drawdown year (2002)	3.46%
Best Profitability year (2007)	85.19%

Table 4: Performance of a long/short fund based on buying the top decile and shorting the bottom decile stocks

The cumulative return of NAV 100 HKD at the beginning of Year 2000 would stand at 3535.80 by Aug'2011! The graph of the cumulative NAV (assuming no transaction cost, slippages, cost of capital etc.) is as shown in Fig 2.

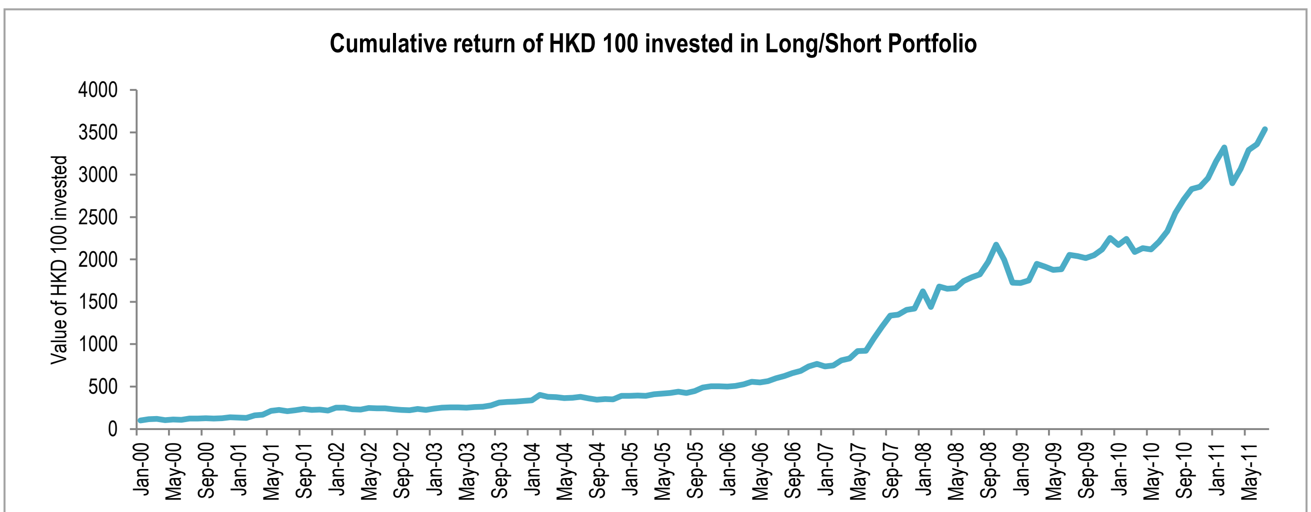

Fig2: Cumulative value of HKD 100 invested in Jan 1999 in a Long/Short portfolio constructed using the unified factor model in HK market and re-balanced on monthly basis (assuming no transaction costs)

As evident from Table 4, the Long/Short portfolio would have made a whopping 36% annualized return over the last 10 years (assuming no slippages, transaction costs and cost of capital) with an annualized Sharpe of 1.67. More importantly the portfolio would have made money every single year - with worst drawdown year being +3.46% (2002) and the best year being +85.19% (2007).

7. Conclusion

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

Firstly the 32 factors were chosen from a diversified set of mean reversion, momentum and fundamental factors. Data was first broken into in-sample (1999-2008) and out-of-sample (2009-2011) periods. A historical back-test for the last 13 years from 1999 to 2011 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 HK market (both in in-sample and out-of-sample periods) over the last 13 years and is not biased towards a given kind of market environment like bull, bear or range bound market. The top 4 factors which were finally shortlisted were: -0.5*3YrRet_zscore+0.5*90DayRet (Mean Reversion Factor), SlopeMonthly_zscore (Momentum Factor), EV/EBIDTA_zscore (Valuation Factor) and ROE_zscore (Growth Factor).

The monte-carlo simulation was then carried out to come up with a single unified factor. The final result of monte-carlo simulation showed that if we give an equal weightage of 25% to each of the 4 factors the overall unified factor is very stable in long term. Hence showing that in the long run stocks in HK market is not biased towards value, growth, momentum or mean reversion.

The overall results of the factor model shows that the top decile stocks outperform the bottom decile stocks with an average of 2.48% monthly and an annualized Sharpe ratio of 1.13. A fund constructed out of being long the top decile stock and shorting the bottom decile stock in equal dollar value would have yielded an average annualized return of a whopping 36% (without any transaction cost, slippages and cost of capital) and an annualized Sharpe of 1.67 with all 11 years from 2000 to 2011 yielding positive returns.

Overall it can be concluded that using techniques of sound historical back-testing, monte-carlo simulation and filtering from a universe of 30+ factors to 4 factors has yielded significantly improved and diversified factors which has stayed stable and generated consistent alpha in the HK market over the last 13 years.

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