Spring Valley
Asset Management

Research Series

The average is better than average

July 2019

Contact Details

+1 212-592-9300
info@springvalleyam.com

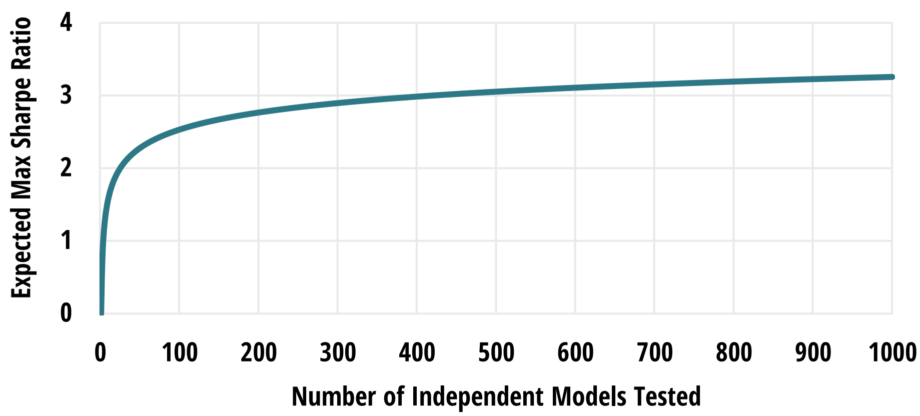
Contents

1. Introduction

Researchers often devote a significant amount of time trying to determine the optimal, or best performing, configuration of a trading model. With the proliferation of data and advances in high-performance computing, it is trivial to optimize millions, even billions, of trading models and parameter sets. While these developments are undoubtedly powerful, researchers are virtually guaranteed to find something that performs very well despite not having any useful predictive ability. It merely fits idiosyncrasies, or noise, in the underlying dataset. In machine learning, this is called overfitting. Indeed, overfitting is now believed to be responsible for the failure of discoveries made in empirical finance to deliver in practice [1, 2].

Exhibit 1

Expected maximum Sharpe ratio after multiple tests



Source: Spring Valley Asset Management, Bailey and Lopez De Prado [1]

Using the approach of Bailey and Lopez De Prado [3], Exhibit 1 demonstrates how easy it is to overfit by simply testing enough variations. For example, after testing only ten independent models, a researcher is expected to find a strategy with a Sharpe ratio of 1.6, even if the true Sharpe ratio is zero.

A classic example of this effect is an investment manager sending out a weekly recommendation to go long or short the stock market [4]. Starting with 100,000 prospects, the manager recommends 50% of them to go long and the other 50% to go short. If the stock market goes up, another recommendation is sent to the 50,000 prospects that were told to go long in the prior week. Again, half are recommended to go long, and the other half to go short. After ten weeks, 97 people will have received 10 consecutive correct stock market recommendations. Since the probability of that occurring is very small ($0.5^{10} = 0.00098$), the prospects would mistakenly believe the manager has an edge.

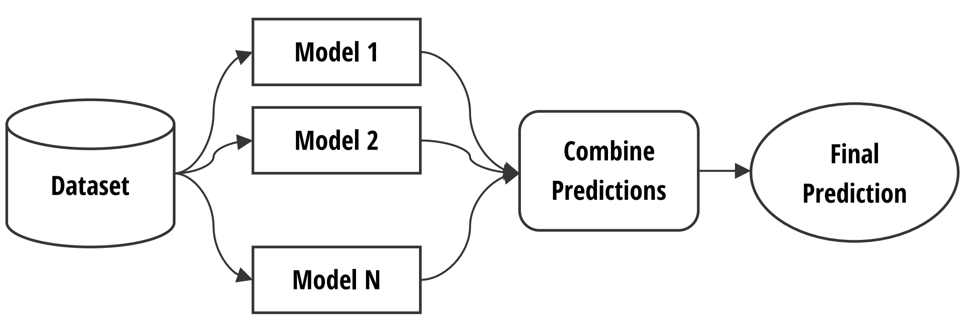
However, researchers are not typically dealing with completely independent models. They will often churn an idea using different, but related approaches. For example, testing simple and exponential moving averages to determine the trend and varying the parameters. This focus on optimizing an individual model lessens its ability to generalize and overlooks an important diversification opportunity.

Since expected returns and volatilities are unknown quantities and must be estimated from historical data, differences in performance between models can merely be the result of estimation error. Therefore, the expected Sharpe ratio across all models is simply the average. Any deviation in performance from this average is likely random and becomes an uncompensated source of risk. If this is the case, it is advantageous to combine the models. In machine learning, this is called an ensemble.

Ensembles pool the predictions from many different models. If the models are imperfectly correlated, the combination can result in superior predictive power.

Exhibit 2

Ensemble illustration



Source: Spring Valley Asset Management. For illustrative purposes only.

Ensembles are similar in principle to the wisdom of crowds. As James Surowiecki [5] succinctly describes the concept:

> *"If you ask a large enough group of diverse, independent people to make a prediction or estimate a probability, and then average those estimates, the errors each of them makes in coming up with an answer will cancel themselves out. Each person's guess, you might say, has two components: information and error. Subtract the error, and you're left with the information."*

An early observation of this effect was in 1906. The English scientist Francis Galton came across a contest to guess the weight of an ox. Galton obtained the results from almost 800 people who participated. No individual estimated the weight exactly. However, the average across all submissions, 1,197 pounds, was remarkably close to the actual weight of 1,198 pounds. Galton called this *vox populi* or "the voice of the people."

If we fast forward to today, we find this concept being applied with algorithms as evidenced by the Netflix Challenge where the first team to achieve a 10% improvement over Netflix's internal algorithm received a $1 million prize. The team from AT&T Labs, *BellKor*, led the competition after one year. Their single best model achieved a 6.58% improvement. However, when they combined all their models, 107 in total, into an ensemble, they realized a gain of 8.43%. They ultimately merged with two other teams to win the Netflix Challenge. In their submission, they noted:

> *"Predictive accuracy is substantially improved when blending multiple predictors. Our experience is that most efforts should be concentrated on deriving substantially different approaches, rather than refining a single technique. Consequently, our solution is an ensemble of many methods."* [6]

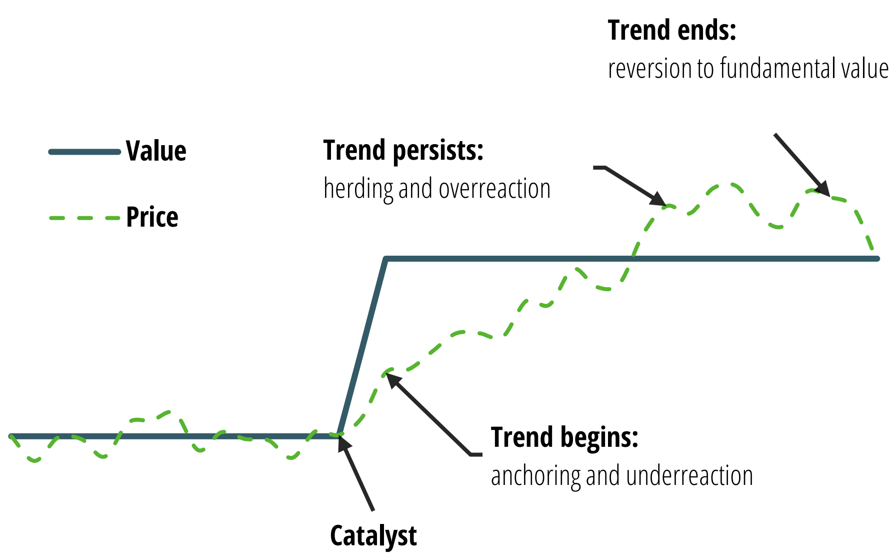
Ironically, the team that came in second place was named *The Ensemble*.

2. Methodology

Our focus will be exclusively on trend following. It has been researched extensively in academia and applied by hedge funds for decades [7, 8]. In its simplest form, the trend is determined using the return over some recent period. For example, if the return on the S&P 500 over the past 12 months is positive, the investor goes long the S&P 500. If the return over the past 12 months is negative, the investor sells short the S&P 500.

Exhibit 3

Trend following illustration



Source: Spring Valley Asset Management. For illustrative purposes only.

There are many explanations for its efficacy. One example is rooted in behavioral economics [9]. It theorizes that the lifecycle of a trend is a manifestation of investor under- and overreaction. If prices initially underreact to new information, prices are slow to fully incorporate changes in fundamental value. However, trends can persist beyond fundamental value as investors overreact and herding takes over. Trends ultimately reverse as a deviation from fundamental value cannot go on indefinitely.

A simple approach to exploiting price trends is time series momentum (TSMOM). Every 21 days, we take the sign of the 12-month return and then scale the signal to target 10 percent volatility. The return for each market s at time $t+1$ is therefore:

$$r_{r,t+1}^{TSMOM,s} = sign\left(r_{t-12,t}^{s}\right)\frac{10\%}{\sigma_t^s} r_{t,t+1}^{s}$$

where $r_{t-12,t}^{s}$ is the 12-month return and σ_t^s is an exponentially weighted moving average of volatility over the prior 60 days. The return of a diversified portfolio is simply the average return across all markets:

$$r_{r,t+1}^{TSMOM} = \frac{1}{S_t}\sum_{s=1}^{S_t} sign\left(r_{t-12,t}^{s}\right)\frac{10\%}{\sigma_t^s} r_{t,t+1}^{s}$$

where S_t is the number of markets available at time t. Each diversified portfolio is then scaled to 10% volatility using the prior 36 months of returns. This ensures that the portfolios target a consistent amount of risk over time.

Now that we have established our baseline approach to trend following, we need a metric to evaluate performance. One of the most common statistics used for this purpose is the Sharpe ratio, $\widehat{SR}$, which represents the amount of return per unit of volatility:

$$\widehat{SR} = \frac{\hat{\mu} - R_f}{\hat{\sigma}}\sqrt{q}$$

where $\hat{\mu}$ is the sample mean return, $\hat{\sigma}$ is the sample standard deviation of returns, R_f is the risk-free rate, and q is the annualization factor.

However, Sharpe ratios calculated from historical data are subject to estimation error. That is, given a Sharpe ratio and the number of observations, we can calculate a range of values that can contain the true Sharpe ratio. We use the approach in Lo [10] to determine the 95% confidence interval for the Sharpe ratio:

$$\widehat{SR} \pm 1.96 \times \sqrt{\left(1 + \frac{1}{2}\widehat{SR}^2\right) / T}$$

where T is the number of observations.

We acknowledge the flaws in using the Sharpe ratio and assuming a normal distribution. We use these methods for simplicity, and our results are qualitatively unchanged by our assumptions.

Also, for the sake of demonstration, we have ignored the role of important practical issues such as transaction costs, slippage, and taxes.

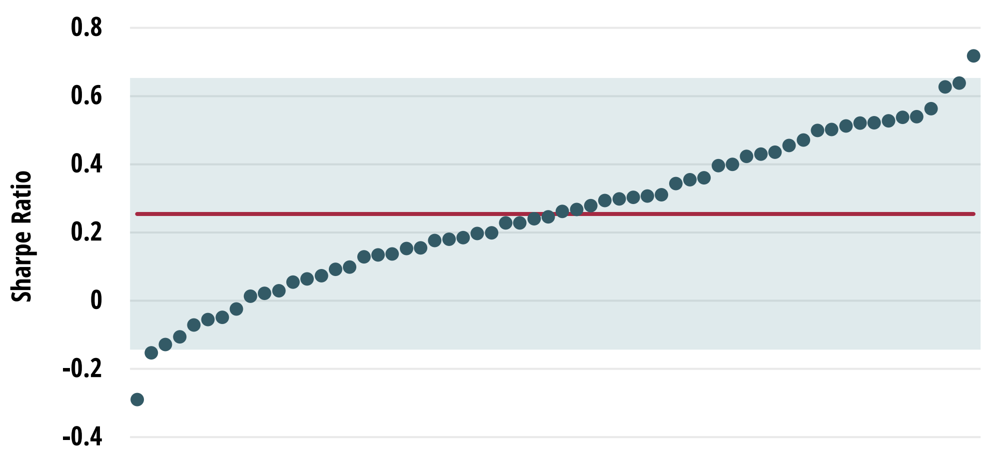
3. Universe Selection

The universe we select to test our theory can have a tremendous impact on our analysis. Therefore, our universe should be economically intuitive and have broad representation across asset classes. Attempting to optimize our universe for the past is futile and will diminish our ability to generalize the future. We will apply our TSMOM signal on a sample of 60 futures markets – 14 bond markets, 13 equity indices, 9 currency pairs, and 24 commodities.

Exhibit 4 plots the sorted Sharpe ratios for each market using TSMOM. We also show the average Sharpe ratio and the 95% confidence interval around that average.

Exhibit 4

Sorted Sharpe ratios for each market using TSMOM and 95% confidence interval



Source: Spring Valley Asset Management

We observe that 57 out of the 60 markets fall within the 95% confidence interval. This means we cannot statistically validate that any of the 57 markets are better for trend following than another. Though three markets fall outside of the 95% confidence interval – Cocoa, Kansas Wheat, and Euro Bunds – there is nothing intuitive to suggest such markets are materially better or worse at exhibiting trending behavior relative to the others. We also come to the same conclusion at the asset class level.

Exhibit 5

Performance by asset class using TSMOM

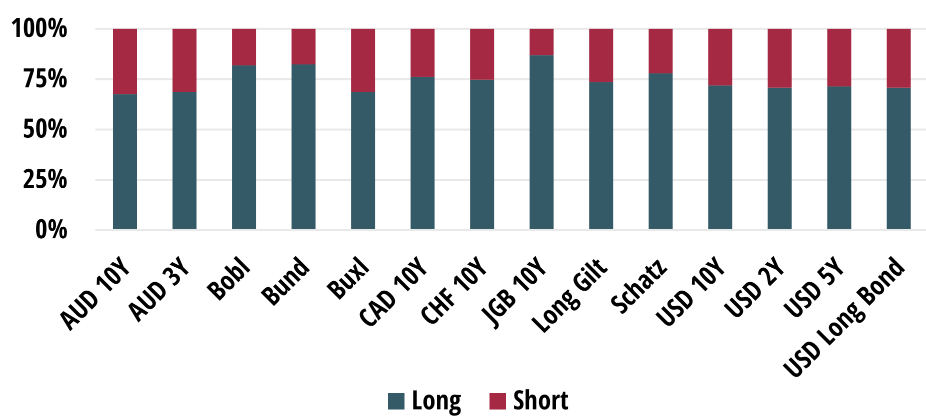
	Bonds	Equity Indices	Commodities	Currencies
Mean	6.01%	5.73%	6.21%	5.37%
Volatility	10.32%	10.43%	10.74%	10.57%
Sharpe	0.58	0.55	0.58	0.51
T	2.87	2.71	2.85	2.51

Source: Spring Valley Asset Management

It is also essential to understand sample bias. For example, the asset class with the highest t-statistic was bonds. However, our sample is characterized by a secular decline in interest rates. Exhibit 6 displays the percentage of time our strategy was long our universe of bonds. On average, the portfolio was long 75% of the time. This does not undermine trend following's applicability to bonds. It merely provides context to what would have been concluded to be the best performing asset class.

Exhibit 6

Percentage of time long and short bonds

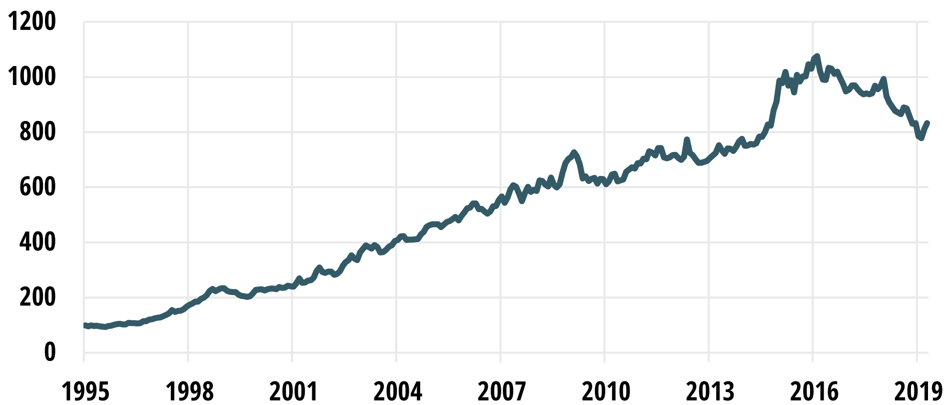


Source: Spring Valley Asset Management

We then aggregate all markets to create a diversified portfolio. The average Sharpe ratio across markets is .25, and the average across asset classes is .56. However, the diversified portfolio realizes a Sharpe ratio of .89. This demonstrates the power of combining many markets that are imperfectly correlated.

Exhibit 7

Diversified portfolio across all markets



	TSMOM
Mean	9.28%
Volatility	10.42%
Sharpe	0.89
T	4.39

Source: Spring Valley Asset Management

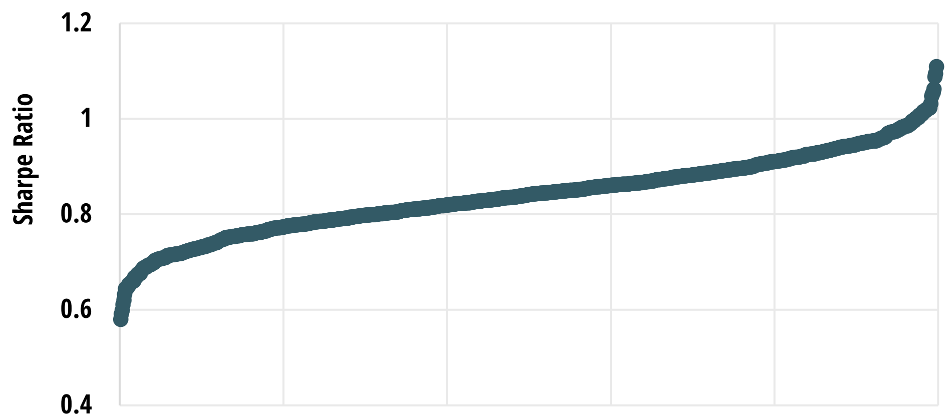
If a researcher attempted to prove that any individual market or asset class was superior to another, not only would they overestimate the expected out-of-sample performance of those markets, but they would overlook a significant diversification opportunity.

To further demonstrate the risks of universe selection, we will randomly choose a subset of 45 markets from our original 60 and create a diversified TSMOM portfolio. We will repeat this procedure 1,000 times.

Exhibit 8 displays the dispersion in Sharpe ratios. The 95^{th} and 5^{th} percentiles are .97 and .71, respectively. Since we are targeting 10% volatility, this results in annualized returns of 9.7% and 7.1%. For context, this would result in an 18% total return differential over five years. Also, with 95% confidence, the true Sharpe ratio is between .44 and 1.24. Since all universe combinations are within this range, any outperformance between them cannot be relied upon in the future.

Exhibit 8

Sorted Sharpe ratios of diversified TSMOM strategy with randomly selected markets



Source: Spring Valley Asset Management

4. Defining a Trend

There are many equally valid ways to determine the trend. Another approach is to use moving averages. Typically, an investor will buy(sell) when the current price, or a short horizon moving average, crosses above(below) a longer horizon moving average. The most common moving averages are simple and exponential. A simple moving average of prices is computed as:

$$SMA_t(n) = \frac{1}{n}\sum_{i=0}^{n-1} P_{t-i}$$

where n is the number of periods. We then create a signal $SMACROSS_t^s(n)$ for each market using the sign of the difference between the current price P_t^s and the n period simple moving average:

$$SMACROSS_t^s(n) = sign(P_t^s - SMA_t^s(n))\frac{10\%}{\sigma_t^s}$$

Alternatively, we can use the exponential moving average EMA_t:

$$EMA_t = (1-\alpha)EMA_{t-1} + \alpha P_t$$
$$EMA_0 = P_0$$

where α is the decay parameter. We will then construct a signal $EMACROSS_t^s$ using the sign of the difference between the short horizon, EMA_s, and long horizon, EMA_l, exponential moving averages:

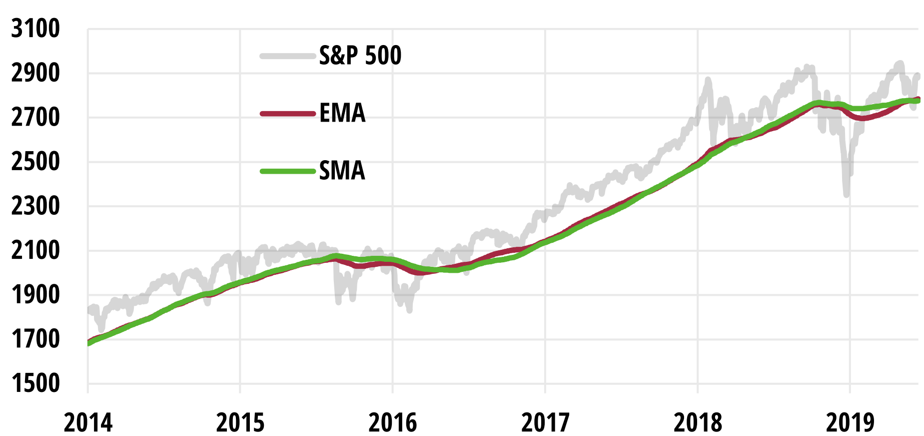
$$EMACROSS_t^s = sign(EMA_s - EMA_l)\frac{10\%}{\sigma_t^s}$$

where the EMA_s is one quarter the length of EMA_l.

Mathematically, our three definitions are highly related [11, 12]. For example, the lag associated with an n period simple moving average is $(n-1)/2$, whereas the lag of an exponential moving average is $\alpha/(1-\alpha)$. Therefore, an n period simple moving can be approximated by an exponential moving average with $\alpha = (n-1)/(n+1)$. Although imperfect, Exhibit 9 reveals how closely they track each other. It is also evident that the exponential moving average is more responsive due to the higher weight on more recent observations.

Exhibit 9

Simple vs. exponential moving average (n=200)

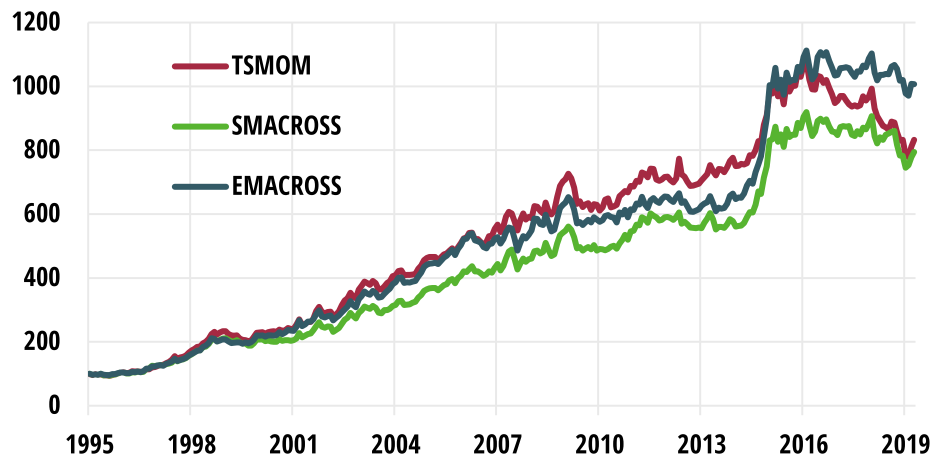


Source: Spring Valley Asset Management

Exhibit 10 shows the performance for all three of our definitions at a horizon of 12 months. Statistically, the Sharpe ratios are indistinguishable from one another. The average Sharpe ratio is .91, and the standard error is .21, which means that with 95% confidence, the true Sharpe ratio is between 1.31 and 0.51. The Sharpe ratios for all definitions fall within this interval.

Exhibit 10

Performance across definitions



	TSMOM	SMACROSS	EMACROSS
Mean	9.28%	9.09%	10.07%
Volatility	10.42%	10.48%	10.41%
Sharpe	0.89	0.87	0.97
T	4.39	4.28	4.77

Source: Spring Valley Asset Management

The evidence suggests we should not favor any one definition over another. In addition, there are substantial risks in choosing any single definition. Exhibit 11 shows that the return differential between the best and worst performing definitions over rolling twelve-month periods can be substantial, reaching up to 27%. However, the outperformance of any one definition is the result of randomness and is unlikely to repeat itself in the future.

Exhibit 11

Rolling 12-month return differential between best and worst performing definition

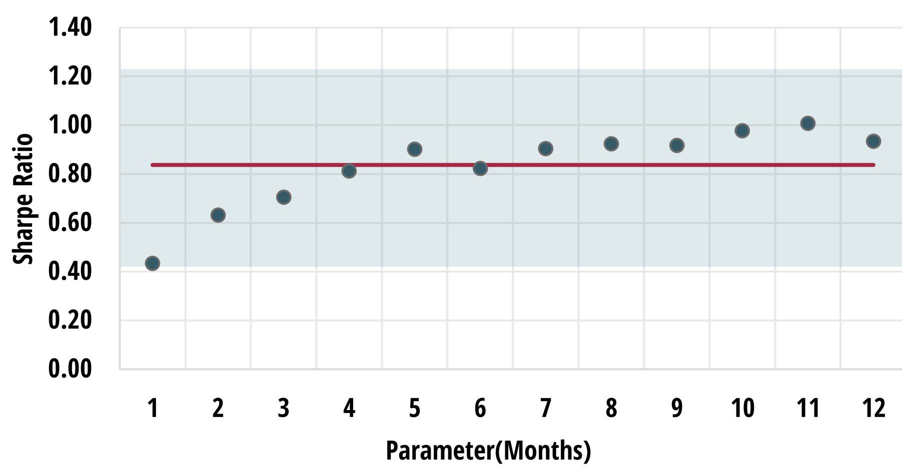
Source: Spring Valley Asset Management

5. Parameter Selection

We have been using information over the prior 12 months to determine whether a given market was trending higher or lower. However, the literature has found between 1 and 12 months to be statistically and economically significant [13]. There lacks a clear economic rationale for which window is "best." Therefore, we will take the average across our three definitions using each parameter between 1 and 12 months for a total of 36 strategies.

Exhibit 12

Sharpe ratios across parameters and 95% confidence interval

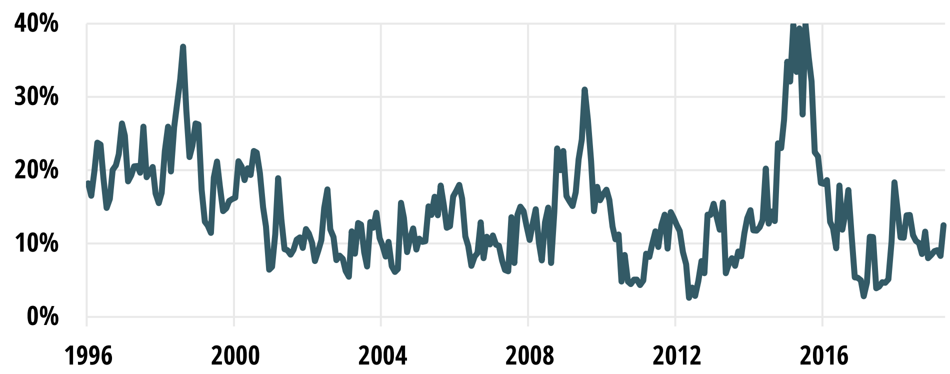


Source: Spring Valley Asset Management

The Sharpe ratios in Exhibit 12 are statistically indistinguishable from one another. However, again, in Exhibit 13, the short-term differentials between parameters can be tremendous, reaching up to 40% over 12 months.

Exhibit 13

Rolling 12-month return differential between best and worst performing parameter



Source: Spring Valley Asset Management

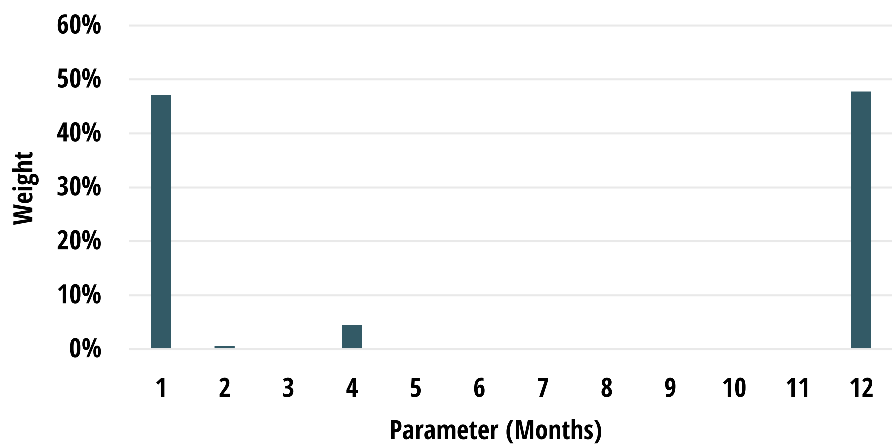
We also see that 1-month is easily the worst performing parameter and would be the most obvious candidate for exclusion if trying to determine what is optimal. However, it has the least amount of information overlap with the other parameters. For example, the 1-month parameter represents 1/6th of the information in the 6-month parameter. Whereas the 3-month parameter represents 50% of the information in the 6-month parameter.

To demonstrate the point, using the approach in [14], we determine the weights to each parameter that would result in the most diversified portfolio over our historical sample.

Consistent with our intuition, Exhibit 14 shows that combining the 1-month and 12-month parameters resulted in the greatest diversification. This confirms the “barbell” weighting revealed in [15]. However, the point here is not to determine the optimal combination of parameters. It is to show that optimizing for a specific parameter not only leads to overstated expectations but neglects additional opportunities for diversification.

Exhibit 14

Weights across parameters for the most diversified portfolio



Source: Spring Valley Asset Management

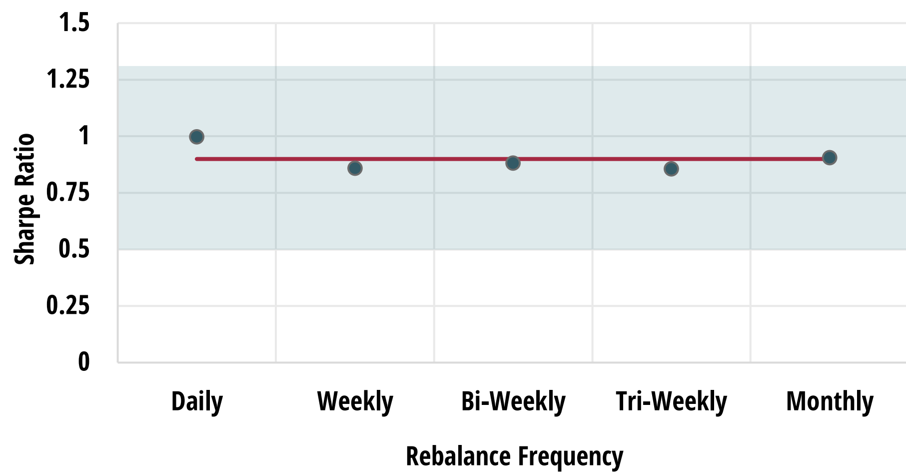
6. How often do we rebalance?

Now that we have analyzed 36 equally valid ways to measure and exploit a trend (3 definitions across 12 parameters), we need to examine how often we should rebalance. Rebalancing is fundamental in maintaining a diversified portfolio. In addition, by analyzing different rebalancing frequencies, a researcher can understand the decay of a signal, or how far into the future a signal is predictive. For example, a high-frequency signal could be ineffective if rebalanced monthly. Conversely, it can be inefficient to rebalance a low-frequency signal every day.

Up to this point, we have been rebalancing every 21 days, a monthly rebalancing frequency. Therefore, we examine rebalancing every 1, 5, 10, 15, and 21 days, or daily, weekly, bi-weekly, tri-weekly, and monthly. Each rebalancing frequency will be the average of 36 strategies (3 definitions across 12 parameters) for a total of 180 strategies.

Exhibit 15

Sharpe ratios by rebalance frequency and 95% confidence interval



Source: Spring Valley Asset Management

Here we have broken out each rebalancing frequency, and again we find that there is no statistical evidence that any rebalancing frequency is preferable over another. However, when we look at performance over short periods, we observe considerable variation in performance, reaching up to 13% over twelve months.

Exhibit 16

Rolling 12-month performance differential between best and worst performing rebalance frequency

Source: Spring Valley Asset Management

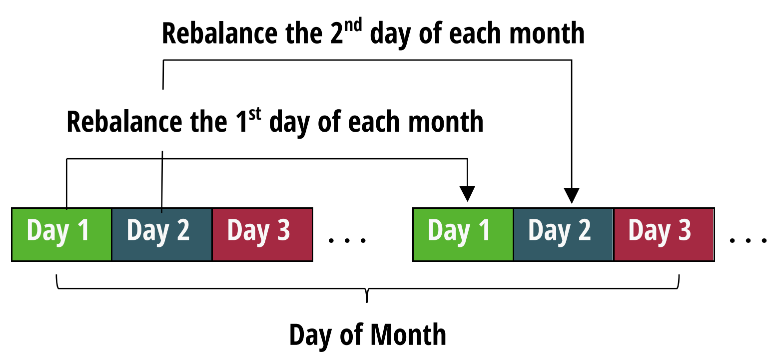
7. When do we rebalance?

The date on which the portfolio is rebalanced is an overlooked piece of the research process. However, portfolios constructed with identical methodologies but rebalanced on different days can exhibit significantly different return profiles over time.

For example, consider a strategy that rebalances monthly. The portfolio can be rebalanced on the first day of every month, the last day of every month, or any day in between. There are approximately 21 days in a month, which means there are 21 possible ways to rebalance monthly. Ultimately, the choice of when to rebalance introduces path dependency, also known as timing luck [16].

Exhibit 17

Monthly rebalancing alternatives

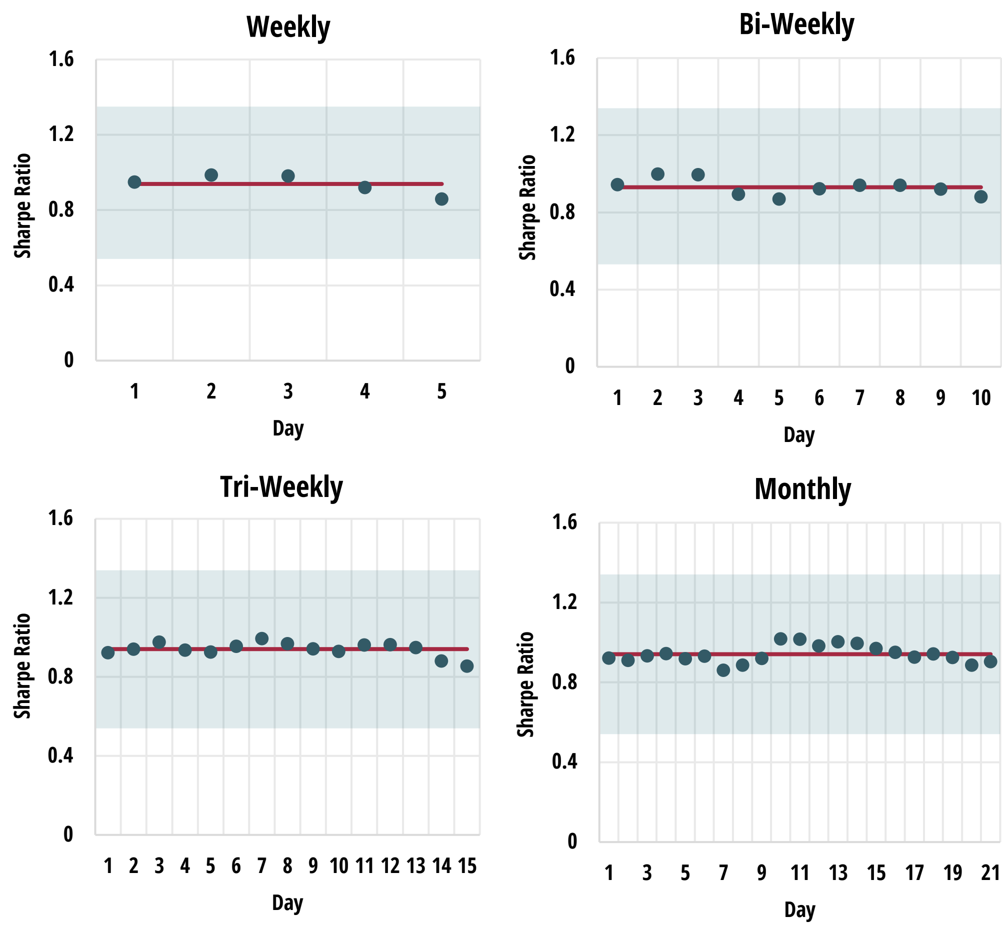


Source: Spring Valley Asset Management. For illustrative purposes only.

Since we are using daily data, a daily rebalancing frequency does not exhibit any path dependency (although it does in practice when considering intraday data), we show the performance differentials for each date within each rebalancing frequency. Within each date is the average of 36 strategies (3 definitions across 12 parameters) for a total of 1872 strategies (36 strategies across 52 days). Exhibit 18 plots the Sharpe ratios for each date within each rebalancing frequency.

Exhibit 18

Rebalancing date within each rebalancing frequency and 95% confidence interval

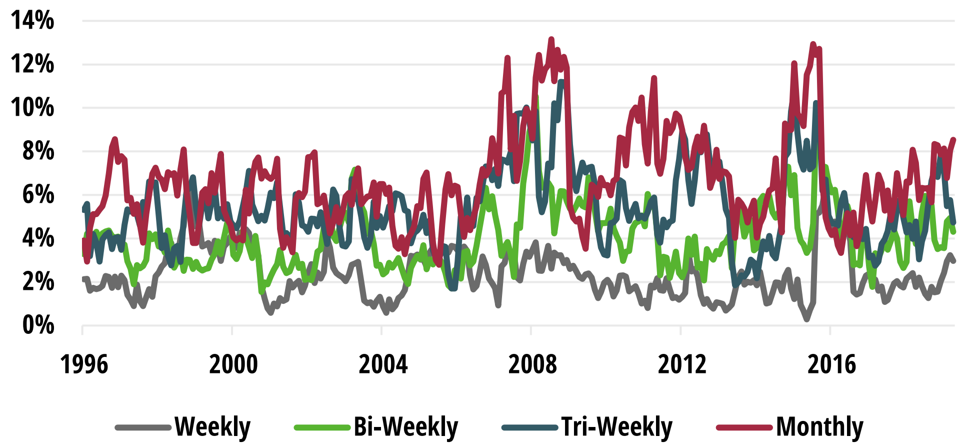


Source: Spring Valley Asset Management.

Exhibit 19 plots the rolling twelve-month performance differentials between the best and worst performing rebalance dates within each rebalancing frequency. We observe the greatest dispersion within the monthly rebalancing frequency. This is logical considering the broader range of alternatives to rebalancing monthly. Overall, however, this proves to be an essential, and potentially critical, detail that deserves much more attention.

Exhibit 19

Rolling 12-month return differential between best and worst performing date within each rebalancing frequency



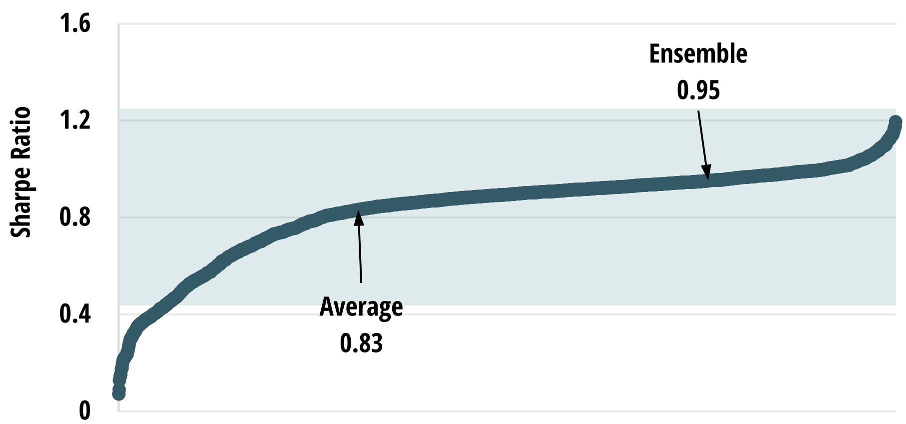
Source: Spring Valley Asset Management

8. Ensemble Performance

Our investigation has yielded 1872 models across several dimensions to substantiate the simple hypothesis that markets exhibit trends. We have also seen that all configurations are legitimate, albeit correlated, ways of exploiting this idea. As a result, our best estimate for the out-of-sample Sharpe ratio of any individual model is simply the average across all

Exhibit 20

Sorted Sharpe ratios across all 1872 models and 95% confidence interval



Source: Spring Valley Asset Management

variations. However, as shown in Exhibit 20, the average Sharpe ratio is .83, but the ensemble realizes a Sharpe ratio of .95, which puts it in the 75th percentile of all variations.

We can understand this result using Grinold's Fundamental Law of Active Management [15, 17] which states that a manager's Information Ratio (or in our case, the Sharpe ratio), IR, is the result of two factors: skill and breadth:

$$IR = IC \times \sqrt{N}$$

where IC is the information coefficient, skill, and N is the number of independent investment opportunities, breadth. For example, consider two portfolio managers with the same investment skill. However, one manager can apply their skill across a broader set of opportunities. This manager has a higher level of breadth and is expected to outperform the other manager despite having the same investment skill.

Since any individual model represents one independent investment opportunity, we can understand the improvement of an ensemble by computing its breadth. For this purpose, we will use the Diversification Ratio [14], DR:

$$DR = \frac{\sum_{i=1}^{N} w_i \sigma_i}{\sigma_P}$$

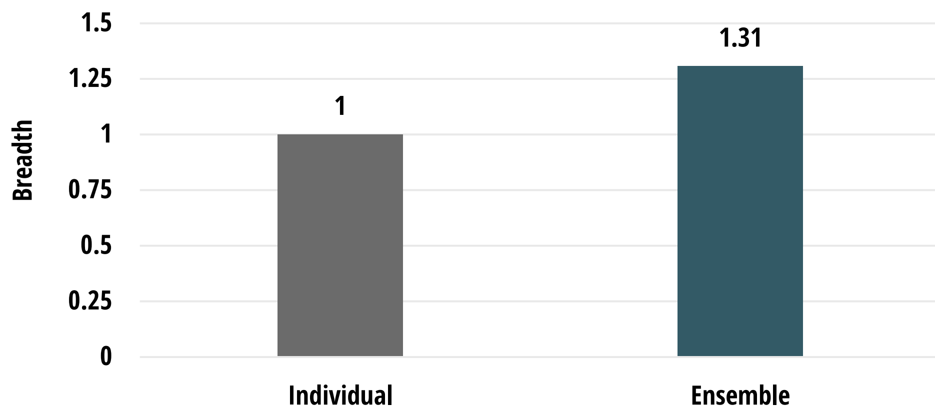
where $\sum_{i=1}^{N} w_i \sigma_i$ is the weighted average volatility, in our case equally weighted, across all models and σ_P is the volatility of the ensemble. The number of independent opportunities is simply the diversification ratio squared, DR^2.

The ensemble has 1.31 independent opportunities as opposed to 1 for any individual model. While this number does not appear to be significant, it results in a 14% increase in the Sharpe ratio.

We also would like to acknowledge the specifications that lie below the confidence intervals. Of the 112 that are outside of the lower boundary, 100 of them belong to the 1-month parameter. We found this to be, in general, the worst performing parameter. However, not only did we find the 1-month parameter to be one of the most diversifying, but we have no theory to suggest why it would not be valid.

Exhibit 21

Number of independent investment opportunities

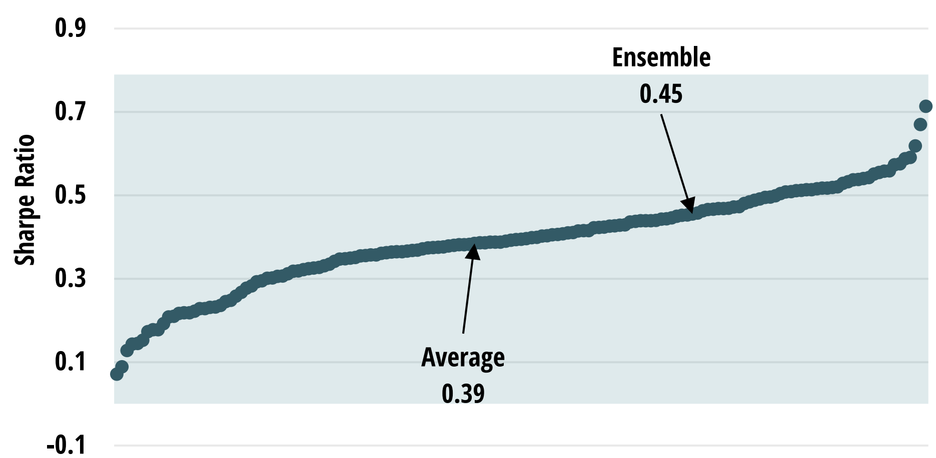


Source: Spring Valley Asset Management

When we aggregate all models with a 1-month parameter, we see that internally they are statistically indistinguishable from each other. Therefore, our best guess for the performance of any individual model using a 1-month parameter is simply the average of .39. However, as shown in Exhibit 22, an ensemble of models using the 1-month parameter generates a Sharpe ratio of .45. Again, this demonstrates the effectiveness of an ensemble.

Exhibit 22

Sorted Sharpe ratios for all models with 1-month parameter and 95% confidence interval

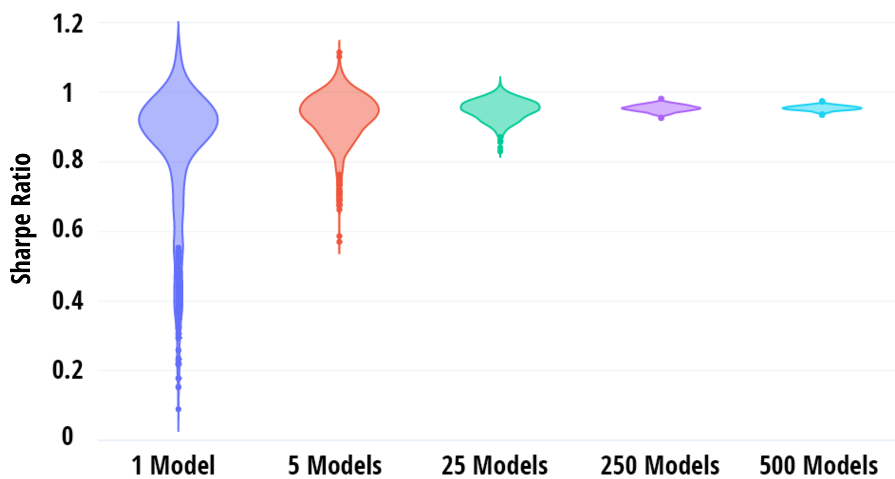


Source: Spring Valley Asset Management

We've also considered deviating from the average portfolio to be an uncompensated source of risk. A simple way to demonstrate how we have minimized this risk is to show how the distribution of Sharpe ratios is affected by the number of models contained in the ensemble. To do this, we take 1000 combinations of n randomly selected models from our universe of 1872. Exhibit 23 shows the results for n of 1, 5, 25, 250, and 500 models. It is evident that the distribution of Sharpe ratios narrows by combining a larger number of models and therefore reduces the risk of deviating from the average portfolio.

Exhibit 23

Distribution of Sharpe ratios for randomly sampled model combinations of size n



9. Importance of Economic Theory

> *"If data is all that matters, then we cannot derive assumptions from theory, and without assumptions, quantitative tools cannot be used, hence data is useless."* [18]

The forces of arbitrage in financial markets substantially reduces the signal-to-noise ratio. As a result, it is much more susceptible to overfitting relative to the natural sciences. However, unlike physicists, we are analyzing a human construct. We understand the underlying dynamics of financial markets and can create logical theories using this knowledge. Having an economic theory lessens the chance that a discovery is spurious and allows us to derive assumptions from which we can determine the appropriate quantitative methods to use. Also, theory helps distinguish between what happened in the past and the uncertainty of the future. It is possible that a strategy created using historical data was real given the market conditions of the sample [19]. However, if a different set of conditions present themselves out-of-sample, these strategies can fail. We can create reasonable theories about our strategies in the context of conditions unknown to our sample. Therefore, a good check on our propensity to overfit is building on a solid economic foundation that is general in applicability.

10. Conclusion

Financial time series are noisy. With the proliferation of data and high-performance computing, it is easier than ever to confound noise with signal. Without considering theory and the distribution of tested variations, researchers will be disappointed to learn that what is optimal in-sample is simply the result of randomness. Therefore, our best forecast for all variations out-of-sample is the average, and any deviation from that average is an uncompensated source of risk. The best approach, then, is to construct an ensemble of all variations. Since they are imperfectly correlated, the performance of the ensemble can outperform, and minimize the risk of deviating from, the average variation.

Appendix

Markets and Data Sources

All historical data is from Bloomberg.

Commodities

Aluminum, Natural Gas, Crude Oil, WTI Crude, Heating Oil, Coffee, Corn, Cotton, Copper, Cocoa, Soybeans, Sugar, Live Cattle, Lead, Gasoil, Nickel, Kansas Wheat, Wheat, Lean Hogs, Feeder Cattle, Gold, Silver, Zinc, and Unleaded Gasoline.

Fixed Income

Euro Schatz, Euro Bobl, Euro Bund, Euro Buxl, Long Gilt, Canada 10-year bond, Swiss 10-year bond, Japan 10-year bond (TSE), Australia 3-year bond, Australia 10-year bond, U.S. 2-year note, U.S. 5-year note, U.S. 10-year note, and U.S. long bond.

Equity Indices

Nikkei 225, SPI 200, FTSE/MIB, S&P/TSX 60, CAC 40, SMI, Hang Seng, AEX, IBEX 35, OMXS30, FTSE 100, DAX, and S&P 500.

Currencies

Japanese Yen, Canadian Dollar, British Pound, Euro (German mark prior), New Zealand Dollar, Australian Dollar, Swiss Franc, Swedish Krona, and Norwegian Krone.

References

[1] C. R. Harvey, Y. Liu, C. Zhu, 2015, *...and the Cross-Section of Expected Returns*, Available at SSRN: https://ssrn.com/abstract=2249314

[2] D. H. Bailey, J. Borwein, M. López de Prado, Q. J. Zhu, 2015, *Pseudo-Mathematics and Financial Charlatanism: The Effects of Backtest Overfitting on Out-of-Sample Performance*, Notices of the American Mathematical Society 61(5), 458-471.

[3] D. H. Bailey, M. López de Prado, 2014, *The Deflated Sharpe Ratio: Correcting for Selection Bias, Backtest Overfitting and Non-Normality*, Journal of Portfolio Management 40(5), 94-107.

[4] C. R. Harvey, Y. Liu, 2014, *Evaluating Trading Strategies*, Journal of Portfolio Management 40(5), 108-118.

[5] J. Surowiecki, 2005, *The Wisdom of Crowds: Why the Many are Smarter Than the Few*, Abacus.

[6] R. M. Bell, Y. Koren, C. Volinsky, 2007, *The BellKor solution to the Netflix Prize*, https://www.netflixprize.com/assets/ProgressPrize2007_KorBell.pdf

[7] T.J. Moskowitz, Y. H. Ooi, L. H. Pedersen, 2011, *Time series momentum*, Journal of Financial Economics 104(2), 228-250.

[8] B. Hurst, Y. H. Ooi, L. H. Pedersen, 2013, *Demystifying Managed Futures*, Journal of Investment Management 11(3), 42-58.

[9] A. Tversky, D. Kahneman, 1974, *Judgement under uncertainty: heuristics and biases*, Science 185, 1124-1131.

[10] A. W. Lo, 2002, *The Statistics of Sharpe Ratios*, Financial Analysts Journal 58(4), 36-52.

[11] V. Zakamulin, 2016, *Market Timing with Moving Averages: Anatomy and Performance of Trading Rules*, Available at SSRN: https://ssrn.com/abstract=2585056

[12] A. Levine, L. Pedersen, 2016, *Which Trend Is Your Friend?*, Financial Analysts Journal 72(3), 51-66.

[13] B. Hurst, Y. H. Ooi, L. Pedersen, 2017, *A Century of Evidence on Trend-Following Investing*, Available at SSRN: https://ssrn.com/abstract=2993026

[14] Y. Choueifaty, T. Froidure, J. Reynier, 2013, *Properties of the Most Diversified Portfolio*, Journal of Investment Strategies 2, 49-70.

[15] C. Hoffstein, 2018, *Measuring Process Diversification in Trend Following*, https://blog.thinknewfound.com/2018/07/measuring-process-diversification-in-trend-following/

[16] C. Hoffstein, J. Sibears, N. Faber, 2018, *Rebalance Timing Luck: The Difference Between Hired and Fired*, Available at SSRN: https://ssrn.com/abstract=3319045

[17] R. C. Grinold, 1989, *The Fundamental Law of Active Management*, Journal of Portfolio Management 15(3), 30-38.

[18] D.H. Bailey, 2018, *The Problem With Financial Oracles*, https://mathinvestor.org/2018/11/the-problem-with-financial-oracles

[19] M. Harris, 2016, *Limitations of Quantitative Claims About Trading Strategy Evaluation*, Available at SSRN: https://ssrn.com/abstract=2810170

Disclaimer

The information set forth has been obtained or derived from sources believed by the authors and Spring Valley Asset Management, LLC ("SVAM") to be reliable. However the authors and SVAM do not make any representation or warranty, express or implied, as to the information's accuracy or completeness, nor does SVAM recommend that the attached information serve as the basis of any investment decision. This document has been provided to you for information purposes and does not constitute an offer or solicitation of an offer, or any advice or recommendation, to purchase any securities or other financial instruments, and may not be construed as such. This document is intended exclusively for the use of the person to whom it has been delivered by SVAM and it is not to be reproduced or redistributed to any other person. This document has been prepared solely for information purposes. The information contained herein is only current as of the date indicated and may be superseded by subsequent market events or for other reasons. Charts and graphs provided herein are for illustrative purposes only. Nothing contained herein constitutes investment, legal tax or other advice nor is it to be relied on in making an investment or other decision.

There can be no assurance that an investment strategy will be successful. Historic market trends are not reliable indicators of actual future market behavior or future performance of any particular investment which may differ materially and should not be relied upon as such.

The information in this document may contain projections or other forward-looking statements regarding future events, targets, forecasts or expectations regarding the strategies described herein, and is only current as of the date indicated. There is no assurance that such events or targets will be achieved and may be significantly different from that shown here. The information in this presentation, including statements concerning financial market trends, is based on current market conditions, which will fluctuate and may be superseded by subsequent market events or for other reasons.

The investment strategy and themes discussed herein may be unsuitable for investors depending on their specific investment objectives and financial situation. Please note that changes in the rate of exchange of a currency may affect the value, price or income of an investment adversely.

Neither SVAM nor the authors assume any duty to, nor undertakes to update forward looking statements. No representation or warranty, express or implied, is made or given by or on behalf of SVAM, the authors or any other person as to the accuracy and completeness or fairness of the information contained in this presentation, and no responsibility or liability is accepted for any such information. By accepting this document in its entirety, the recipient acknowledges its understanding and acceptance of the foregoing statement.

There is no guarantee, express or implied, that long-term return and/or volatility targets will be achieved. Realized returns and/or volatility may come in higher or lower than expected. PAST PERFORMANCE IS NOT AN INDICATION OF FUTURE PERFORMANCE.

Simulated performance results (e.g., quantitative backtests) have many inherent limitations, some of which, but not all, are described herein. No representation is being made that any fund or account will or is likely to achieve profits or losses similar to those shown herein. In fact, there are frequently sharp differences between hypothetical performance results and the actual results subsequently realized by any particular trading program. One of the limitations of simulated results is that they are generally prepared with the benefit of hindsight. In addition, simulated trading does not involve financial risk, and no hypothetical trading record can completely account for the impact of financial risk in actual trading. For example, the ability to withstand losses or adhere to a particular trading program in spite of trading losses are material points which can adversely affect actual trading results. The simulated results contained herein represent the application of the quantitative models as currently in effect on the date first written above and there can be no assurance that the models will remain the same in the future or that an application of the current models in the future will produce similar results because the relevant market and economic conditions that prevailed during the hypothetical performance period will not necessarily recur. There are numerous other factors related to the markets in general or to the implementation of any specific trading program which cannot be fully accounted for in the preparation of hypothetical performance results, all of which can adversely affect actual trading results. Discounting factors may be applied to reduce suspected anomalies. This backtest's return, for this period, may vary depending on the date it is run. Simulated performance results are presented for illustrative purposes only.

There is a risk of substantial loss associated with trading commodities, futures, options, derivatives and other financial instruments. Before trading, investors should carefully consider their financial position and risk tolerance to determine if the proposed trading style is appropriate. Investors should realize that when trading futures, commodities, options, derivatives and other financial instruments one could lose the full balance of their account. It is also possible to lose more than the initial deposit when trading derivatives or using leverage. All funds committed to such a trading strategy should be purely risk capital.