Systematising Start-Trade and Strategy Upgrade Decisions

In an effort to improve our strategies we upgrade them to incorporate the latest research periodically, generally between once and twice per year. Each time there is an upgrade, the firm must decide when to implement the change to the trading program. At issue is the question of whether there exists an optimal time to adopt an upgraded strategy.

Is there a time when the "old" strategy is due for a loss, while the upgraded one is ready to surge? Is it better to upgrade at new equity highs or during a drawdown? Is there a way to systematically avoid "double-dip drawdowns" whereby the old strategy is in a drawdown, an upgrade is made to a strategy not in drawdown, and the upgrade then proceeds into drawdown? Is it possible to systematically capture "double-dip run-ups," whereby we exit the old strategy at new highs, just in time to participate in a recovery from drawdown in the upgraded strategy?

There are related questions concerning the timing of new account initiations. When a client wishes to open an account to participate in one of our programs, is there a more intelligent start-trade method than to begin trading as soon as the account is funded and the paperwork is complete? Perhaps new accounts should wait for some amount of drawdown before beginning to trade? Perhaps they should initiate only at new equity highs? Perhaps they should dollar-cost-average into the account?

This study seeks to answer these questions and others by examining both the statistical properties of our program returns as well as some implications of ideas axiomatic to behavioural finance.

Difficulties in Assessing Short-Run Performance In general, the difficulty of judging whether a trading program is performing to expectations increases as the length of the performance sample decreases. To couch it in engineering terms, this is a consequence of trading results consisting of both "signal" (the positive edge our systems extract from the markets) and "noise" (fluctuations indistinguishable from random that occur interspersed with the system alpha). As the length of the performance sample increases, the alpha generated by the system (signal) adds up over time, while the random fluctuations (noise) tend to cancel one another out. It follows that the time necessary to verify that sample performance is representative of signal rather than noise becomes less as the signal-to-noise ratio increases and greater as the signal-to-noise ratio decreases.

What's the Frequency, Kenneth? A good analogy is a radio broadcast with varying levels of static interfering with the clarity of the transmission, depending on our distance from the broadcast tower. When we are very close to the tower the signal is strong, and we don't have to listen for very long to determine whether someone is communicating on that channel. Any sensible transmission stands out loud and clear against the background of static noise. As we approach the very edge of broadcast range, it may take hours of careful listening to distinguish any meaningful signal from the background noise, even if the signal is just a simple pulse wave, but especially if it is a spoken message or other complex communication. There may be lots blips or bleeps that might be a part of a message, but only after tedious verification may we decode the whole message by bits and pieces.

Perhaps it is S-O-S…

Also to keep in mind, in view of our analogy, is that in the case of trading strategies we may need to view the quality of the message itself with some suspicion, without regard to S/N ratio. After all, trading strategies are only models that produce signal rather than noise on an average basis, historically. We already know that because market behaviour is notoriously hard to predict accurately, our strategy is likely to broadcast erroneous (but strong) signals part of the time.

Whereas we may be reasonably sure that a weather broadcaster has checked with reliable sources before telling us that the temperature is 71 degrees F, our analogous trading strategy may report that the temperature is likely to be (based on all historically available information) 73 degrees F, plus or minus 15 degrees standard deviation. The second broadcast contains much less usable information than the first. We don't know the temperature outside for

sure: It could be 50 degrees or 90. Neither is surprising based on the available information. Alas, it is the best our models may achieve so far.

Carrying forward the radio analogy, our trading systems produce signal / noise (Sharpe or Information) ratios of about 4-to-1 in the best cases, whereas an audio broadcast at reasonable FM radio quality has an S/N ratio of better than 70-to-1 (high fidelity audio may reach S/N ratios greater than 100-to-1). In other words, if alpha were a radio broadcast, we would be pretty far from the tower, straining hard to hear its tune, which hopefully is not merely a siren song. This being the case, we are generally much more confident in our ability to estimate relatively long term performance expectations (3 years +) as opposed to very short term ones (3 months and less). In the short term we expect much more of our performance to be the result of random fluctuation than in the long term. The question is, how much more?

The answer may be found by calculating the Sharpe ratio for periods other than the industry standard, an annualized ratio based on monthly periods. A daily Sharpe ratio hints at the amount of signal we observe per unit of noise on a daily timeframe. The monthly Sharpe ratio does the same for monthly periods, as does the annual Sharpe ratio for yearly periods. If all is right with the world, we should see that S/N (Sharpe) ratios increase as periodicity lengthens. Our strategies should extract more profit per year per unit of annual volatility than profit per day per unit of daily volatility. By analysing the relationship between long term and short term S/N ratios, we may be able to surmise the minimum reasonable period for meaningful performance evaluation.

Sharpe Ratios for Multiple Periodicities The Sharpe ratio, or more accurately, the information ratio (since this version does not include a term for the risk free rate) represents the compound rate of return divided by the standard deviation of price changes for any periodicity desired. It is common to annualize returns and standard deviations using monthly return data as the source to generate an annualized Sharpe ratio. However, we may calculate a Sharpe ratio for any unit of time down to the smallest time unit for which we have periodic return data. Since we track returns for our strategies all the way down to the daily time frame, we may begin measuring Sharpe ratios here.

A daily Sharpe ratio consists of the daily compound growth rate of the series divided by the standard deviation of daily returns. As the graph below indicates, using daily data as the source, the daily Sharpe ratio is .17 (noise is almost 6 times greater than signal) while over a 5 year time horizon, the Sharpe is 28.77 (the signal is almost 29 times as great as noise). The Sharpe ratio increases steadily on a logarithmic scale as periods grow longer because the returns the strategy generates are cumulative and grow geometrically over time, while the standard deviation of returns does not grow geometrically. In other words, as time goes on the S/N ratio goes up, and more of our return may be attributed to our strategy, less to random noise.

The next graph shows a similar picture, but this time monthly returns are used as the data source. We see that our monthly Sharpe ratio is .68, while our annual Sharpe is 3.02. Dividing the former into the latter, we note 4.4 times more volatility per unit of return in the monthly returns than in the annual returns. For periods longer than a year, the Sharpe ratios are high enough to give us some degree of confidence that returns over those periods are representative of strategy return, not random variability.

However, for time frames of less than 1 year it is not clear that periodic returns are distinguishable from random. As noted, monthly periods have about 4.5 times as much variability attributable to chance as annual periods, and daily periods have about 4.5 times as much variability attributable to chance as monthly periods.

This evidence confirms the hypothesis that S/N ratios, and thus the ability to infer meaning from periodic returns, increase with the length of the period measured. It also quantifies the amount of variability inherent to shorter time frames compared to the longer timeframes (1-5 years) that we acknowledge as statistically significant. In short, we have to be willing to trust a number 4.5 times more variable than annual return estimates in order to extract meaning from individual monthly returns, and we have to be willing to trust a number 20 times as variable as annual return estimates to extract meaning from daily returns. Even quarterly returns are 2.5 times as variable as annual returns.

With this in mind, together with evidence from previous work indicating that there is not an extreme degree of autocorrelation in the daily or monthly returns of our programs, we may do well to approach any short term timing of our strategy with caution if the goal is to avoid the pitfalls that can accompany a strategy upgrade. Until we have strategies with monthly Sharpe ratios of 2-3 or more, the conservative choice is to act as if monthly returns are random.

Implications for Upgrade Timing If we believe the analysis presented above, and we see that monthly performance is extremely difficult to distinguish from random, much less to predict, then we must abandon the notion of timing upgrades based on any type of short term strategy performance. To be clear, there is no evidence that the level of a strategy's future draw-up, draw-down, or daily, weekly, or monthly performance are in any way predictable based on recent past returns.

Note that the previous statement does not mean that there is no value in long term performance measures such as historical drawdown amount, long term growth rates, etc. These types of metrics may still be used to intelligently time a strategy upgrade. For example, it is not farfetched to say that a strategy near a maximum historical drawdown presents an optimal entry point. If historical performance holds, the strategy is likely to come out of drawdown and reach new highs. In the case of an upgrade strategy near historic lows, and an existing strategy at new highs after a large run-up, the decision is a no-brainer… as long as we believe the new strategy is superior to the old. However, this would be a rare situation, and is likely not very useful in the typical upgrade scenario. It simply does not make sense to wait, perhaps for years, for everything to line up just right before upgrading if we have a superior strategy waiting in the wings. Moreover, upgrades are generally similar enough to the existing strategy that the upgraded strategy will not be doing poorly during a time that the existing strategy is doing very well.

So, we must come up with a different method of timing the upgrade. Research suggests the answer is simple: If an upgraded strategy is superior to the existing strategy, implement it as soon as possible. The logic of this hopefully needs no further explanation.

Timing New Account Initiations: A Different Animal On its face, it may seem that our analysis of upgrade timing would carry over to account initiation. If we can't time putting up new accounts based on recent past performance, shouldn't we just gain exposure to our profitable trading strategy as soon as possible? The answer is yes… for an investor that views his initial capital no differently from trading profits. Experience shows us that such an investor is atypical. Most investors are more risk averse concerning their initial capital investment than they are concerning the growth on that capital. That is, they are more willing to give back gains and less willing to experience drawdown on initial capital. This bias is well documented in behavioural finance and is just one instance where a client may forgo optimal risk adjusted returns in preference of minimizing perceived risk on initial capital.

Summary

Accounting for this bias, there is a set of circumstances under which timing account initiation basis current performance is called for:

1. The manager and client both assume that historical performance is a reasonable guide to future performance (i.e. a strategy with a 15% maximum historical drawdown is likely to bottom out at roughly that same drawdown level in the future).

2. The client is more averse to losing initial capital than to missing a run of good performance.

When these two criteria are met, a client's needs are best accommodated by waiting to initiate the strategy until a drawdown of sufficient magnitude occurs. Because of our assumption that past performance is indicative of future results, an investment near a historic drawdown level is more likely to be followed by a return to new highs than by further drawdown.

Note that for any positive expectation strategy, the investor is likely to make less absolute return by using any timing method other than putting the account up as soon as possible. This is because, for a strategy that makes money,

performance in any future period is more likely to be positive than negative. That is, a client is more likely to miss positive performance than negative by waiting to initiate an investment. This is why our second assumption above is critical. In reality, it is the rare investor that will not second guess an account initiation timing decision if he ends up waiting out a nice run-up. That is human nature.

In view of the facts, perhaps the best service we may provide as money managers is to communicate the pros and cons of different start-trade methods to clients prior to investment. An assessment may be made of the client's preferences and an appropriate start-trade method chosen. There are really only three possibilities:

1. Put up the account immediately. This suits the investor that wants to receive the maximum benefit of the strategy and is not sensitive to drawdowns on initial capital.

2. Put up the account only when a drawdown of sufficient magnitude occurs. This suits the investor that wishes to minimize initial drawdown, perhaps at the expense of missed profits.

3. Dollar cost average. For the investor of sufficient means, funding the account over time diversifies risk among the two previously listed methods.

None of these three methods alter the risk-adjusted returns of the strategy, but when dealing with investor preferences risk-adjusted returns are not always of primary consideration.

On a final note, as managers that are paid management and incentive fees, there is a clear preference for start-trade method from our perspective. Given that we like to stabilize our fee base, method #2 is our preference. New money coming in during losing periods mitigates the effect of drawdowns on assets under management. It also ensures that coming out of the drawdown high-water marks are at relatively low points for the new accounts funded during drawdown. If we did not have a preference for stabilizing assets, method #1 would be optimal for the same reasons it is optimal for a client that wishes to make as much money as possible.

Conclusion

• Analysis of Sharpe ratios calculated over various performance periods indicates that short-run performance of less than about 1 year is so noisy as to be indistinguishable from random.

• Because of low signal-to-noise ratios on short time frames, timing upgrades based on recent performance is likely more art than science, with the results of such timing more likely to be attributable to chance than skill.

• In view of these facts, upgrades should be put in place as soon as we are confident that the upgrade is a better strategy than the current one.

• In consideration of biases well documented in behavioural finance literature, an investor's needs may be best met by waiting for a drawdown to incept a new account