LOMBARD ODIER
INVESTMENT MANAGERS

Investment viewpoint

Toning down the optimism of backtests

For professional investor use only • Alternatives

October 2022

Alexey Medvedev
Senior Portfolio Manager,
Systematic Alternatives & Equities

Laurent Joué
Senior Portfolio Manager,
Systematic Alternatives & Equities

Marc Pellaud
Senior Portfolio Manager,
Systematic Alternatives & Equities

Cesar Vaurof Moura
Senior Analyst, Systematic
Alternatives & Equities

Key points

- Alternative risk premia offer valuable diversification from traditional asset classes. However, backtests tend to offer an overly optimistic view of the potential future performance of such strategies
- Performance bias originating from the backtest optimisation can be evaluated numerically for a more accurate assessment of expected returns
- Sharpe ratios can be trimmed to be more accurate; however, as some strategies will survive through luck, due diligence is required

Backtests are wrong, but useful

Past performance does not guarantee future returns. This statement is true of any strategy, be it of a systematic or discretionary nature. Systematic strategies seemingly provide a little more comfort in the form of lengthy historical backtests, which indeed never look bad. What escapes our eyes is the amount of effort spent on optimising these strategies to work best in the past; this does not necessarily mean that their performance will continue to be as strong in the future.

In reality, the more attractive the past performance, the more likely it is that the strategy is "over-optimised"; as a result, its future returns are likely to disappoint. Complex strategies requiring numerous parameters for calibration are obviously more prone to this phenomenon. However, even simple strategies such as those using single-equity factors are not exempt. For example, our review of the industry revealed as many as 35 different metrics used to define the Quality factor. The numerous combinations of these metrics obviously create plenty of room for experimentation.

While backtests are always wrong, they are still useful. Historical returns allow us to better understand when the strategy works and assess its contribution to overall portfolio risk. However, performance tends to be inflated in backtests, which makes it more difficult to accurately predict future portfolio returns. One way to get a more realistic picture is to focus on strategies which have sufficiently long live track records that are consistent with their simulated performance. Of course, the main drawback of this approach is that we will never invest in new ideas.

In this piece of analysis we propose an alternative approach which involves correcting the "over-optimisation" bias inherent in historical backtests. Starting with a comprehensive database of nearly 1400 quantitative risk premia strategies offered by top providers,

Please note: this document focuses on backtests and is intended for information only, not for sales and marketing purposes.

FIG 1. SUMMARY OF THE ALTERNATIVE PREMIA DATABASE

	EQUITY	COMMODITY	BONDS	FX	CREDIT	MULTI ASSET	ALL
Number of strategies	588	305	192	141	115	48	1389
Number of premia	12	9	10	8	8	7	54
Years of live track record[1]	4.0	4.8	3.6	4.0	2.9	3.6	3.7
Sharpe ratio (backtest)[1]	0.62	1.04	0.67	0.56	0.22	1.03	0.69
Sharpe ratio (live)[1]	-0.04	0.13	0.01	-0.01	-0.27	0.20	0.00

[1] Average across strategies. Source: LOIM. Past performance is not an indicator of future results. For illustrative purposes only.

we first estimate the expected drop in Sharpe ratio once the strategy goes live. We then illustrate how to use these findings to predict future performance more accurately for a portfolio of alternative risk premia.

Live versus backtested returns

We built a database of alternative risk premia strategies offered by top providers as of the end of August 2022. Specifically, we required each strategy to have a Bloomberg ticker and an exact date when the strategy went live. The strategy's performance before that date is considered to be a backtest.

Based on provided descriptions, we classified strategies into 54 different risk premia buckets across five asset classes (Equity, Bonds, FX, Credit and Commodities) plus Multi-asset. Figure 1 provides a summary of the strategies in the database.

The last two rows in the table report separate average Sharpe ratios based on backtest and live periods respectively. Unsurprisingly, the back-tested Sharpe ratios are positive across all asset classes. In contrast, the live Sharpe ratios are rather disappointing, with that of Credit in particular being well below zero.

While it is tempting to use the difference between the two averages as a measure of the optimisation bias, we should exercise caution. For most strategies, their live periods fall within the last few years, which are known to have been particularly challenging for alternative premia overall. The observed gap in performance between backtest and live periods can therefore best be explained as a result of the combined impact of the particular environment and the optimisation bias. In order to evaluate the pure impact of the optimisation bias, a slightly more sophisticated analysis is required.

Isolating the optimisation bias

We started by calculating realised Sharpe ratios over fixed windows before and after the live date for each strategy.[1] For a more accurate comparison of performance, we then converted them into excess Sharpe ratios by subtracting the risk-adjusted returns of each strategy's peers; the peer returns were calculated as a risk-based combination of the returns of all live strategies in our database belonging to the same premia bucket.[2] The difference between excess Sharpe ratios before and after the live date of the strategy provides an estimation for its optimisation bias. Of course, there are many specific factors that can impact this value for each strategy; therefore, to arrive at a more accurate measure we should average these values across all the strategies.

Because of the way we have calculated the optimisation bias, strategies with less than one year of live track record are excluded. This condition, along with the requirement that peer returns are available two years prior to the live date of the strategy, reduces the total size of the sample from 1389 to 872 strategies.

Due to the survivorship effect, we can expect strategies with a longer live track record to have a smaller bias. Indeed, those strategies that did badly during the live period are likely to have been decommissioned over time and will therefore not be present in our database. This natural process of eliminating strategies that fail to meet expectations has its benefits. However, it offers no guarantee that the remaining selection will not disappoint in the future. Providers create and modify numerous quantitative strategies every year, so there is a strong possibility that some strategies will survive the test of time purely by luck.

To evaluate the impact of the "age" of the strategy, we ran a regression of strategies' biases on the length of the period they had been live, while controlling for the fixed effect of each asset class.[3] We established that an extra one year of live track record tends to reduce the magnitude of the bias by 0.07 on average, and this coefficient is statistically significant. Among asset classes, we found that Commodity is most prone to over-optimisation, followed by Equity.

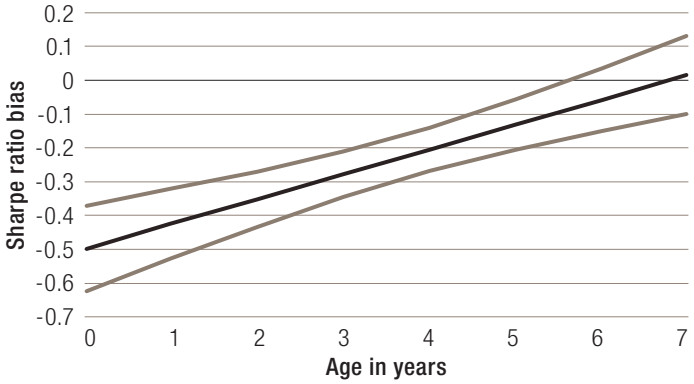
[1] For the backtest returns we used a window of 24 months, while for the live returns we used a shorter window of 12 months to maximise the number of observations.

[2] The returns of peer live strategies are weighted by their inverse daily return volatilities over the previous three months. The benchmark return is further adjusted for its daily volatility over the previous three months, to maintain the constant ex-ante volatility.

[3] Strategies may tend to have different lengths of live track record depending on the asset class. Therefore, controlling for asset classes makes It possible to calculate the pure effect of strategy maturity.

Figure 2 shows the expected optimisation bias as a function of the length of live track record, along with its standard deviation bounds. The magnitude of the optimisation bias at age zero appears to be -0.5. We can interpret this number as the expected drop in Sharpe ratio of a generic strategy after it goes live. As the strategy gets older, the bias decreases as expected due to the survivorship effect. The narrowest standard deviation bounds correspond to the age of the average strategy in our sample (3.7 years). This bias is estimated with the highest confidence, at the level of -0.23 ± 0.06.

FIG 2. OPTIMISATION BIAS IN BACKTESTED RETURNS



Source: LOIM. The middle line is the estimate of the optimisation bias. Standard deviation bounds are indicated by the light-grey curves.

Trimming risk premia

We have combined a number of selected strategies from our database with our in-house implementations to form a diversified portfolio of alternative risk premia (ARP). This portfolio combines 15 different risk premia across equities, bonds, commodities and foreign exchange. To benchmark its performance, we formed a portfolio invested in the same premia with a generic (non-optimised) implementation.[4]

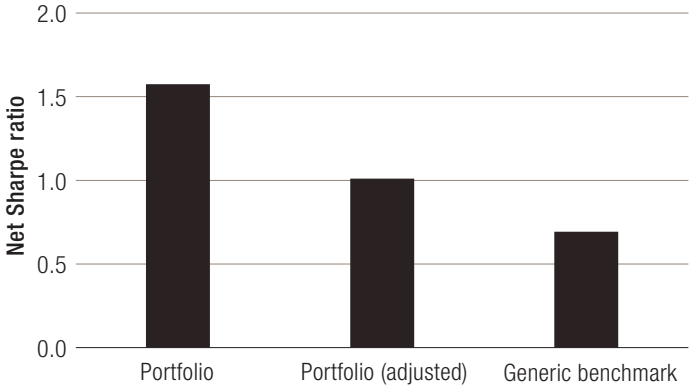
Following our analysis of the optimisation bias, we toned down the risk-adjusted performance[5] of each strategy in the ARP portfolio in the following way:

Adjusted Sharpe Ratio = Sharpe ratio – 0.50 + 0.07 x Years

where Years is the number of years the strategy is live.

Figure 3 shows the net Sharpe ratio of the ARP portfolio (original and adjusted), alongside that of the generic benchmark over the last 10 years. The unadjusted Sharpe ratio is as high as 1.6, which is more than double the risk-adjusted performance of the benchmark. However, after applying an adjustment for the expected optimisation bias, the Sharpe ratio drops to 1.0; this is still above the benchmark but not by a large margin. This result confirms the value added of proprietary implementations of risk premia, while at the same time, providing a more realistic assessment of expected outperformance.

FIG 3. PERFORMANCE OF ARP PORTFOLIO (AUGUST 2012 – AUGUST 2022)[6]



Source: LOIM. Note: Figures are net of costs. Past performance is not an indicator of future results.

Conclusion

Alternative risk premia offer valuable diversification from traditional asset classes. However, forming expectations about their future returns is very difficult, due to the inevitably inflated historical performance of proprietary implementations. In this analysis we have demonstrated how to assess, as far as possible, potential future returns more accurately by accounting for the expected "over-optimisation" bias. This method trims the Sharpe ratios of backtests, imposing larger penalties on strategies with shorter live track records. It should be recognised, however, that since some strategies may survive purely through luck, such a simple solution is far from perfect. With this in mind, a proper due diligence process will still be required to minimise false positives.

[4] A description of a broad set of generic risk premia strategies can be found in "How Do Factor Premia Vary Over Time? A Century of Evidence", by Ilmanen, A. et al. Published in the Journal of Investment Management, Volume 19, No. 4, pp. 15-57 in 2021.

[5] We adjusted strategy returns before live dates using rolling realised volatility.

[6] These performance results are backtested based on an analysis of past market data with the benefit of hindsight, do not reflect the performance of any LOIM product and are being shown for informational purposes only. While the results presented are based on certain assumptions that are believed to reflect actual trading conditions, these assumptions may not include all variables that can affect, or have affected in the past, the execution of trades. The hypothetical portfolio results are based on the following assumptions:
1. The hypothetical portfolio record does not include deductions for brokerage commissions, exchange fees, or slippage.
2. It assumes purchase and sale prices believed to be attainable. In actual trading, the prices attained may or may not be the same as the assumed order prices.
3. The portfolio results do not take into account any tax implications arising from the sale or purchase of securities, which in actual trading do have an impact on gains and losses.

IMPORTANT INFORMATION