Stress-Testing a Tactical Allocation Model: Robustness and Improvements

Performance and resilience across different specifications

QUANTSEEKER
SEP 19, 2025 · PAID

Hi there,

Over the past weeks, I've received several thoughtful questions from subscribers about the tactical allocation model I discussed earlier. Many of you asked about robustness: What happens if we change the weighting scheme? What if we adjust the thresholds or modify the defensive set of assets? Others wondered about possible improvements and whether the strategy holds up across variations.

In today's post, I'll take a systematic look at these issues. My main takeaway upfront is that the framework is surprisingly robust. Across weighting choices, signal specification and parameter shifts, the model continues to deliver equity-like returns with far smaller drawdowns than SPY. Still, there are interesting nuances worth highlighting.

Introduction

I won't repeat the full details of the tactical allocation model here. If you'd like a refresher, please see my earlier post where I outlined the mechanics. Instead, this article focuses on robustness: How the model behaves when we vary key assumptio such as weighting schemes, defensive thresholds, lookback windows, and fallback rules. The goal is to stress-test the framework and see whether its performance hol up across different specifications.

Weighting Schemes

The first dimension to test is *how we combine the defensive assets*. In the base versic used rank-based weights as in the original paper, giving more weight to stronger momentum signals. However, signals are noisy, and it's not certain that allocating more to larger but noisy signals is necessarily optimal. Two natural alternatives are worth examining:

- Equal-weighting the defensive set, where each qualifying asset gets the same share.
- Risk-weighting, or inverse-volatility weighting, where more capital is allocated t lower-volatility defensives. I use a rolling 3-month volatility to determine weigh

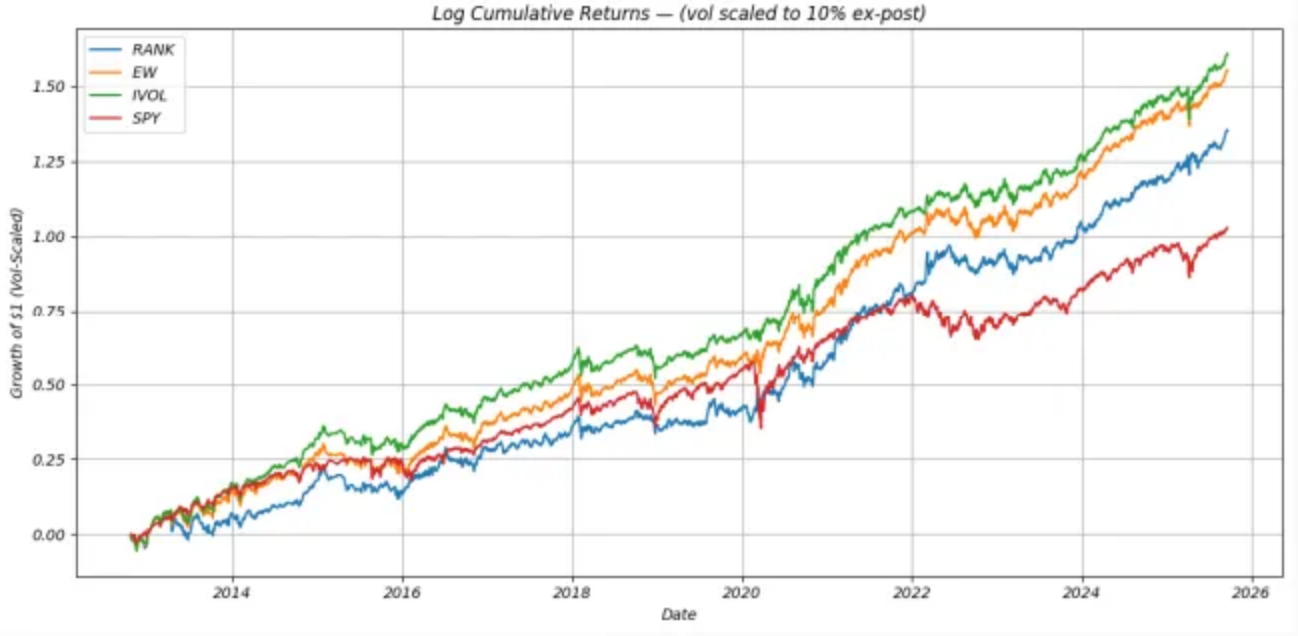
The weighting alternatives only apply to the defensive bucket, whereas the equity exposure, the fallback asset, simply receives the residual weight.

Running the model across rank weights, equal weights, and risk weights yields the following results.

```
=== Summary Stats (vs SPY) ===
              RANK     EW   IVOL    SPY
CAGR %       11.08  12.90  12.87  16.38
Ann.Vol %     8.96   9.02   8.68  17.90
Max DD %     -9.40 -10.10  -9.10 -33.72
Sharpe        1.08   1.26   1.30   0.87
Corr w/ SPY   0.29   0.65   0.65   1.00
DD/Vol       -1.05  -1.12  -1.05  -1.88

=== Annual Returns (%): Strategies vs SPY ===
       RANK     EW   IVOL    SPY
Date
2012  -3.26  -3.13  -2.89  -1.66
2013  10.16  17.95  19.60  33.92
2014  10.28  11.98  13.92  15.61
2015  -3.91  -4.63  -2.44  -0.17
2016  13.94  13.90  14.32  12.00
2017   7.18  11.45  10.94  21.71
2018   0.51  -0.95  -0.57  -4.57
2019   4.62  10.65   9.28  31.22
2020  16.52  18.48  21.15  18.33
2021  22.61  24.01  19.28  28.73
2022  10.13   3.24   4.14 -18.18
2023  11.74  16.32  13.02  26.18
2024  13.61  17.50  17.16  24.89
2025  16.55  15.87  14.61  13.67

=== Average Portfolio Weights (full sample) ===
        TLT    GLD    DBC    UUP   BTAL    SPY
RANK  0.141  0.166  0.151  0.129  0.116  0.297
EW    0.107  0.117  0.104  0.109  0.096  0.469
IVOL  0.091  0.097  0.075  0.191  0.077  0.468
```



When we compare the three weighting schemes, the differences are subtle but meaningful. Equal-weighting and inverse-volatility weighting delivered higher CAGR (≈12.9%) than rank weighting (11%), primarily because they allocate more capital to equities, on average, about 47% versus 30% under rank. That higher equity tilt brings them closer to SPY, with correlations of 0.65 compared to just 0.29 for rank.

Rank weighting, in contrast, produces the lowest equity beta, the smallest max drawdowns, and the most differentiated return stream. Interestingly, equal-weighting and inverse-volatility weighting achieve the highest Sharpe ratios (1.26 and 1.30) compared to 1.08 for rank. In practice, rank weighting may appeal to those explicitly seeking a return stream that has low correlation to equities, while EW and IVOL tend offer higher equity exposure, higher CAGR, and higher Sharpe ratios.

It's comforting to see that equal-weighting, an even simpler construction than rank based weights, produces a higher Sharpe ratio. By spreading capital evenly across the defensives, the model achieves a Sharpe ratio of 1.26, outperforming the original rank based scheme. This suggests that the robustness of the framework does not rely on fine-tuned weightings, as even a naive allocation rule delivers strong risk-adjusted returns.

Excluding Defensive Assets

The main idea of the model is to include defensive assets that are different in nature with low correlation to each other, and which diversify equity risk in different ways. Let's check the performance impact of each specific defensive asset by excluding or asset at a time.

```
=== Summary Stats (Rank-based): Full vs Excluding One Defensive ===
              FULL  EXCL_TLT  EXCL_GLD  EXCL_DBC  EXCL_UUP  EXCL_BTAL    SPY
CAGR %       11.08     12.23     11.31      9.68     10.99      11.67  16.38
Ann.Vol %     8.96      9.50      9.04      9.34     10.56       9.77  17.90
Max DD %     -9.40     -9.07    -12.23    -10.15    -14.53     -13.06 -33.72
Sharpe        1.08      1.14      1.10      0.91      0.93       1.05   0.87
Corr w/ SPY   0.29      0.35      0.27      0.16      0.29       0.45   1.00
DD/Vol       -1.05     -0.96     -1.35     -1.09     -1.38      -1.34  -1.88
```

The exclusion tests confirm that no single defensive asset dominates the strategy, a performance holds up reasonably well even if one is removed. Most exclusions stay close to the full model, though there are nuances:

- Excluding GLD delivers a similar Sharpe as the full model with worse drawdown suggesting gold is a valuable hedge when other assets fail.
- Excluding DBC or UUP weakens results: CAGR falls to 9–10% and Sharpe drops about 0.9. Both provide diversification in specific regimes (inflation shocks for DBC, dollar bull runs for UUP), so their absence seems to hurt resilience.
- Excluding BTAL leads to a bigger drawdown (–13%), confirming its role as an equity hedge during downturns.
- The full mix still looks strongest on balance, offering the best trade-off betweer CAGR, drawdowns, and low correlation to SPY.

Overall, this exercise reinforces the point that robustness comes from *combining heterogeneous defensives* as each asset contributes differently during the sample period.

Transaction Costs

Some readers also asked about transaction costs. These are investor-specific, driven trade size, venue, and execution quality. I assume a one-way cost of 5 bps in the ba case, but I stress-test higher costs. As costs rise toward 25 bps, the strategy's Sharp compresses and converges toward SPY's.

```
=== Summary Stats (Rank-based, varying costs) ===
             RANK_5bps  RANK_10bps  RANK_15bps  RANK_20bps  RANK_25bps     SPY
CAGR %           11.08       10.59       10.11        9.63        9.16   16.38
Ann.Vol %         8.96        8.95        8.95        8.95        8.95   17.90
Max DD %         -9.40       -9.63       -9.86      -10.08      -10.31  -33.72
Sharpe            1.08        1.03        0.99        0.94        0.89    0.87
Corr w/ SPY       0.29        0.29        0.29        0.29        0.29    1.00
DD/Vol           -1.05       -1.08       -1.10       -1.13       -1.15   -1.88
```

Changing Defensive Threshold

While it is intuitive to use the risk-free rate as the hurdle for deciding whether an asset's momentum score justifies an allocation, it is worth asking how sensitive the model is to that choice. To test this, I rerun the strategy with six fixed thresholds of 0 1%, 2%, 3%, 5%, and 7%, and compare them to the benchmark RF-based specificati The results show that performance remains robust: Sharpe ratios and drawdowns ar broadly stable across all thresholds, indicating that the model is not overly depende on the precise choice of hurdle.

```
=== Summary Stats: RF Benchmark and Absolute Hurdles vs SPY (cost = 5 bps) ===
             BENCH   0pct   1pct   2pct   3pct   5pct   7pct    SPY
CAGR %       11.08  10.29  10.66  11.11  11.12  12.36  11.77  16.38
Ann.Vol %     8.96   8.87   8.87   8.99   8.97   9.14   9.68  17.90
Max DD %     -9.40  -9.40  -9.31  -9.99  -9.99  -9.07 -12.29 -33.72
Sharpe        1.08   1.01   1.05   1.08   1.09   1.19   1.07   0.87
Corr w/ SPY   0.29   0.27   0.28   0.31   0.34   0.38   0.49   1.00
DD/Vol       -1.05  -1.06  -1.05  -1.11  -1.11  -0.99  -1.27  -1.88
```

Excluding Lookback Windows

How sensitive is the model to the choice of lookbacks? In its current form, the strate computes each defensive asset's annualized return over 21, 63, 126, and 252 days a

averages the four. Below, I rerun the model while omitting one window at a time.

The results are reassuring: Performance is robust across all exclusions. Sharpe ratios remain broadly similar from spec to spec, but they are consistently lower than the fu specification. That gap suggests the full model benefits from diversifying across horizons, blending short-, medium-, and long-term momentum improves efficiency (higher Sharpe) and reduces drawdowns relative to any single-window subset.

```
=== Summary Stats: Benchmark vs Excluding One Lookback vs SPY (cost = 5 bps) ===
              BENCH  EXCL_21  EXCL_63  EXCL_126  EXCL_252    SPY
CAGR %        11.08    10.43    10.09     10.58     10.41  16.38
Ann.Vol %      8.96     8.85     8.73      8.96      9.14  17.90
Max DD %      -9.40    -8.36   -11.83    -11.00    -12.31 -33.72
Sharpe         1.08     1.03     1.01      1.03      1.00   0.87
Corr w/ SPY    0.29     0.29     0.31      0.31      0.31   1.00
DD/Vol        -1.05    -0.94    -1.36     -1.23     -1.35  -1.88
```

Requiring Fallback Asset to Have Positive Momentum

A thoughtful reader asked: *"What if we only fall back to SPY if it also has positive momentum?"*

In the benchmark setup, the model relies on defensive assets performing well durin equity downturns. But one can imagine a scenario, although not yet observed in the backtest, where all defensive assets exhibit weak momentum *and* SPY is entering a drawdown. In that case, the benchmark would still allocate fully to SPY, which could very costly.

To address this, I rerun the model with an additional filter: SPY must also have a momentum score above the risk-free rate to receive any fallback allocation. If not, t fallback weight instead goes to cash, earning the risk-free rate.

The results show that this modification modestly reduces performance: The Sharpe ratio declines to 0.97 and drawdowns tick higher. However, it adds a layer of protection by ensuring the model never allocates to an asset with poor or negative

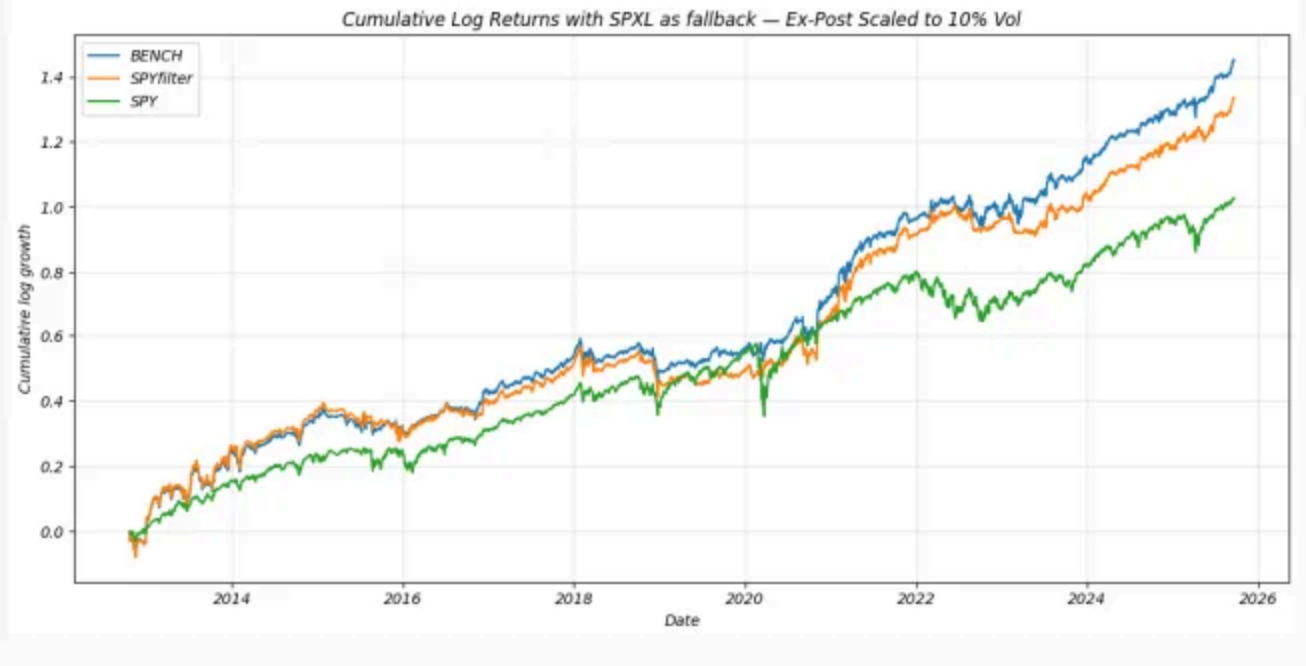
momentum. For more conservative investors, this may be a worthwhile trade-off, offering additional peace of mind even at the cost of some performance.

```
=== Summary Stats: Benchmark vs SPY-filtered fallback vs SPY (cost = 5 bps) ===
             BENCH  SPYfilter    SPY
CAGR %       11.08       9.99  16.38
Ann.Vol %     8.96       8.94  17.90
Max DD %     -9.40     -10.50 -33.72
Sharpe        1.08       0.97   0.87
Corr w/ SPY   0.29       0.16   1.00
DD/Vol       -1.05      -1.17  -1.88
```

If an investor is willing to take on some leverage through a triple-levered ETF, SPXL, the model with a momentum filter on SPXL has offered a CAGR of about 20%, a Sharpe of 1.13, and with a 23% drawdown, all beating SPY over time, as shown belo under *SPYfilter*.

Fallback Asset: SPXL, with momentum filter

```
=== Summary Stats: Benchmark vs SPY-filtered fallback vs SPY (cost = 5 bps) ===
             BENCH  SPYfilter    SPY
CAGR %       23.46      20.27  16.38
Ann.Vol %    17.33      16.47  17.90
Max DD %    -22.00     -23.09 -33.72
Sharpe        1.23       1.13   0.87
Corr w/ SPY   0.61       0.43   1.00
DD/Vol       -1.27      -1.40  -1.88
```



Conclusion and Investor Takeaway

Taken together, these robustness checks show that the tactical allocation model is n fragile. Whether I vary the weighting scheme, exclude individual defensives, raise transaction costs, shift thresholds, drop lookback windows, or require positive momentum for the fallback equity sleeve, the broad picture is the same: Equity-like returns with substantially smaller drawdowns. The nuances matter, equal-weighting boosts Sharpe, gold and BTAL play unique hedging roles, and a SPY momentum filt offers extra downside comfort at the cost of a somewhat lower Sharpe and higher drawdowns, but none of these changes overturn the model's core strengths.

Investor takeaway: This is a flexible, resilient framework rather than a parameter-sensitive and data-mined model. Investors can adapt the specification to their own preferences, whether prioritizing higher Sharpe, lower correlation to equities, or stronger downside protection, while still retaining the core benefits of systematic tactical allocation.

This Substack is reader-supported. To receive new posts and support my work, consider becoming a free or paid subscriber.

Disclaimer: This newsletter is for informational and educational purposes only and should not be construed as investment advice. The author does not endorse or recommend any specific securities or investments. While information is gathered from sources believed to be reliable, there is no guarantee of its accuracy, completeness, or correctness.

This content does not constitute personalized financial, legal, or investment advice and may not be suitable for your individual circumstances. Investing carries risks, and past performance does not guarantee future results. The author and affiliates may hold positions in securities discussed, and these holdings may change at any time without prior notification.

The author is not affiliated with, sponsored by, or endorsed by any of the companies, organizations, or entities mentioned in this newsletter. Any references to specific companies or entities are for informational purposes only.

The brief summaries and descriptions of research papers and articles provided in this newsletter should not be considered definitive or comprehensive representations of the original works. Readers are encouraged to refer to the original sources for complete and authoritative information.

This newsletter may contain links to external websites and resources. The inclusion of these links does not imply endorsement of the content, products, services, or views expressed on these third-party sites. The author is not responsible for the accuracy, legality, or content of external sites or for that of any subsequent links. Users access these links at their own risk.

The author assumes no liability for losses or damages arising from the use of this content. By accessing, reading, or using this newsletter, you acknowledge and agree to the terms outlined in this disclaimer.

3 Likes · 1 Restack

Discussion about this post

Comments Restacks

Write a comment...