Developed Markets Sovereign Bonds Investing: Enhancing Style with Sentiment

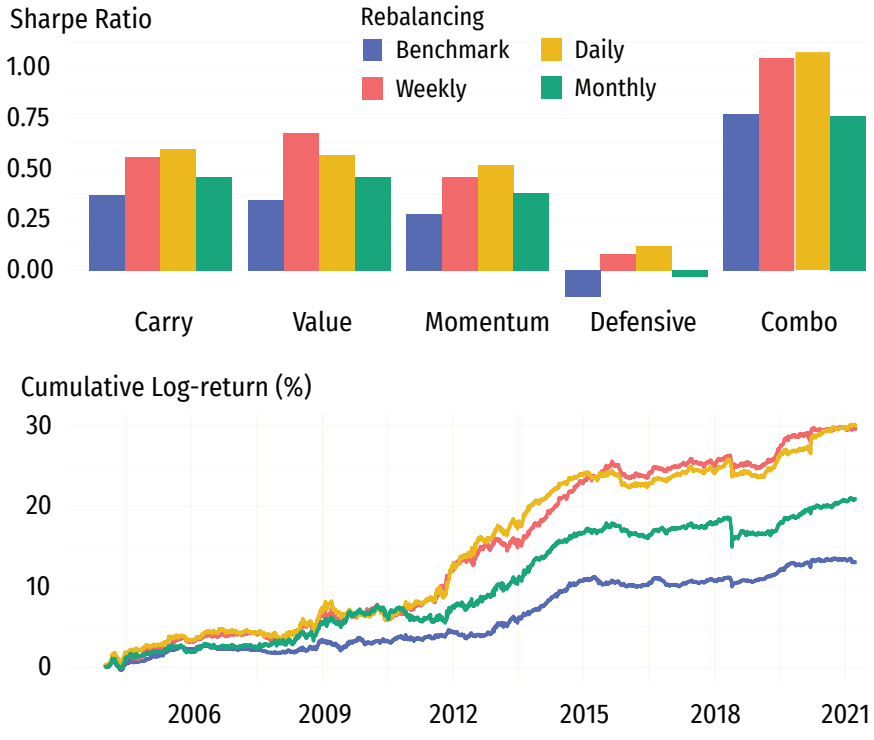
RavenPack
QUANTITATIVE RESEARCH

June 2021

Key Takeaways

With interest rates in developed markets (DM) currently hovering at historic lows, a long-only exposure to DM sovereign fixed income may prove challenging in the years ahead. Style investing offers an alternative way of capturing returns uncorrelated with traditional assets, and has recently been extended to bonds. In this paper, we take one step further and show how standard style factor performance can be enhanced with tilts based on real-time sentiment analytics from RavenPack.

- We construct cross-sectional, dollar and duration neutral long-short portfolios for value, carry, momentum, and defensive styles as our benchmarks, and show that combining them into an equal-weighted strategy results in substantial portfolio diversification benefits.
- We then apply economic news sentiment tilts to benchmark portfolio weights and find that this more dynamic style implementation significantly improves risk-adjusted performance across all four styles, in different country universes and at daily, weekly, and monthly rebalancing frequencies.
- With weekly rebalancing, sentiment boosts the Sharpe ratios of the combined style strategies from 0.77 to 1.08 for the sample of 18 countries, and from 0.43 to 0.69 for the sample of the 6 sovereigns with the most liquid bond futures markets.



Sharpe ratios of sentiment-enhanced styles and cumulative excess returns of combined strategy

Table of Contents

RavenPack Macro Strategy Team

Peter Hafez
Chief Data Scientist

Inna Grinis
Senior Data Scientist

Alan Liu
Data Scientist

research@ravenpack.com
www.ravenpack.com

Americas: +1 646 277-7339
EMEA: +44 (0) 20 3290 7395
APAC: +65 3138 4809

1. Introduction

With sovereign bond yields steadily falling for much of the past several decades **(FIGURE 1)** and interest rates currently hovering near all-time lows, attaining attractive returns in the developed-market fixed income space is likely to be challenging in the years ahead.

Instead of maintaining an outright long exposure, investors may want to access the asset class through style factors. Long-short style investment strategies try to capture distinct risk premia within an asset class without exposing investors to directional risk. They have originally and primarily been applied in the equities space (cf. the seminal work by Fama and French (1992), Jegadeesh and Titman (1993), etc.), and more recently extended to other asset classes, including fixed income (e.g. Asness et al. (2015), Brooks et al. (2018), Brooks and Moskowitz (2017), Fattouche (2018), Gava et al. (2020)).

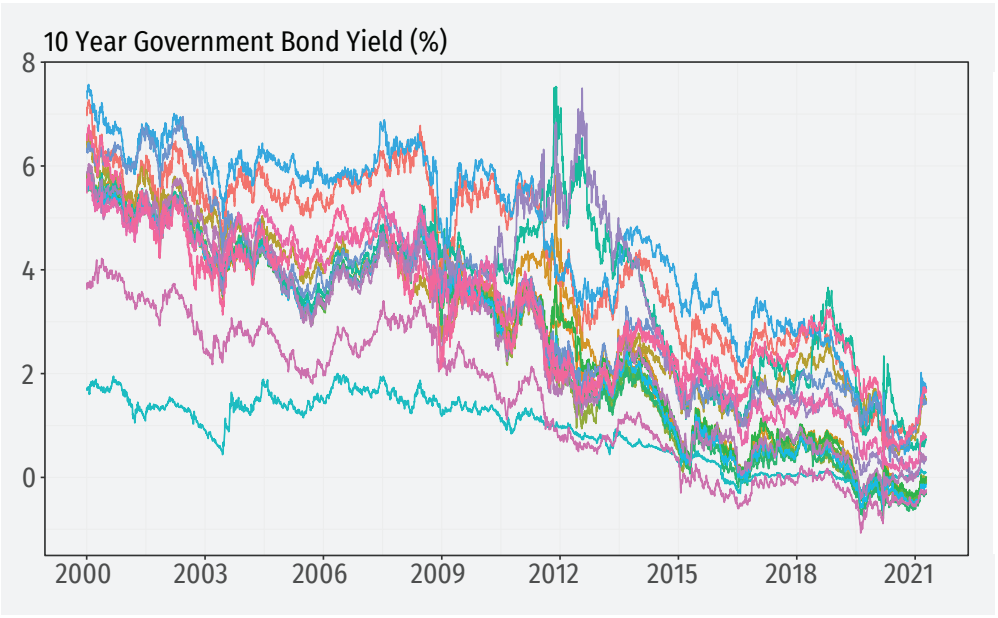
In this paper, we take one step further and show how alternative data can be introduced into value, carry, momentum and defensive sovereign fixed income style investment strategies to enhance performance. We start by constructing established style benchmarks and show that combining them into an equal-weighted strategy results in substantial portfolio diversification benefits. We then apply sentiment tilts based on real-time news sentiment from RavenPack to each style's benchmark portfolio weights. With this more dynamic style implementation, we find that risk-adjusted performance improves significantly across all four factors, as well as in the combined strategy.

In the fixed income space, RavenPack data has proven valuable at corporate (Cai et al. (2017)), municipal (Cheng (2021)), and sovereign (Erlwein-Sayer (2018)) levels in forecasting yield and CDS spreads, spread volatility, etc.[1]

This paper is the first attempt to incorporate RavenPack news sentiment into a systematic fixed income style investing framework. The approach we take is similar to that in our recent FX paper where we show that overlays based on RavenPack macroeconomic news sentiment can significantly enhance the performance of traditional currency style factors (Hafez et al. 2020).

The remainder of the paper is organized as follows: Section 2 presents the data, style definitions, and benchmark construction & performance. In Section 3, we explain how sentiment signals are constructed and overlaid with the benchmark portfolios. Section 4 presents and discusses our results. In Section 5, we provide our conclusions, and the Appendix contains further robustness checks.

FIGURE 1: 10 year government bond yields
Source: Macrobond



1 See RavenPack (2019) for an overview of recent academic papers employing RavenPack data in fixed income applications.

2. Data & Style Definitions

2.1 Sovereign Bond Data

In the DM space, the sovereign fixed income style literature often considers either:

- **6 countries:** Australia, Canada, Germany, Japan, the U.K., and the U.S. in studies that focus on the most liquid markets and use 10-year bond futures (e.g. Asness et al. (2015), Gava et al. (2020), Fattouche (2018))
- **13 countries:** the six markets above, plus Belgium, Denmark, France, Italy, the Netherlands, Spain, and Sweden for studies that rely on the J.P. Morgan Government Bond Index (e.g. Brooks et al. (2018))

In this paper, we use the ICE BofA Government Indices, which track the performance of locally denominated sovereign debt, publicly issued by governments in their domestic markets. Qualifying securities must satisfy a number of criteria to be included, and the indices are market capitalization weighted.[2] The geographical coverage spans beyond the 13-country subset noted above, and we therefore decided to extend it in a way that preserves a universe where interest rate risk is the primary driver of returns. Specifically, we include five additional countries, such that our main sample consists of:[3]

- **18 countries:** Australia, Canada, Germany, Japan, the U.K., the U.S., Belgium, Denmark, France, Italy, the Netherlands, Spain, Sweden, Austria, Finland, New Zealand, Norway, and Switzerland

In addition to this 18-country universe, we test our investment strategies on the 6 and 13 country subsets to verify the robustness of our results and to ensure that our findings remain relevant for the investors who want to implement sentiment-enhanced style strategies on these smaller universes.

ICE BofA Government Indices are available for different maturity buckets, as well as for an "All Maturities" group, which contains all qualifying securities with at least 18 months to final maturity at the time of issuance. We focus on the 7–10 year indices to make our results more directly relevant for investors seeking to implement the strategies in the 10-year bond futures markets, but also test our strategies on the "All Maturities" indices to verify robustness.

While our study primarily focuses on local currency excess returns to avoid confounding fixed income performance with FX movements, we also present the results in USD hedged excess returns for transparency.[4] All returns are gross of transaction costs.

We evaluate all strategies within the 02/01/2004 — 31/03/2021 timeframe.

2 For example, in the case of the US Treasury Index, bonds must have at least one year remaining term to final maturity, a fixed coupon schedule, a minimum amount outstanding of $1 billion, etc. See https://indices.theice.com for more details. A full list of bond index mnemonics used in our study is available on request.

3 We refrain from including Ireland and Portugal, as these countries experienced official bailouts within our study period, and therefore have been subject to significant credit risk. Some styles have to be defined differently when credit risk is a primary driver in addition to interest rate risk, which is why we avoid adding these countries to the DM pool.

4 To compute excess returns, we use the US 3-month Treasury bill rate as the cash rate in the USD hedged case, and the OECD MEI short rates (3-Month or 90-Day Rates & Yields, Interbank Rates) in the LC case. Although the latter are monthly, they have been standardized to ensure cross-country comparability, as 3-Month government bill rates are not available at the daily frequency for the full sample of interest.

2.2 Style Factor Definitions

The main focus of this paper is to test the value of adding sentiment to style investing. We therefore define our factors in the most standard and simple way possible. More sophisticated style constructs could potentially boost factor performance, but may in turn raise data mining concerns. Moreover, whatever the chosen style measure, sentiment can be implemented within a proprietary framework using a similar approach.

We consider the four classic style factors: value, carry, defensive, and momentum, as well as a "Combo" strategy equally invested in all four factors.

Value

Value investing is one of the oldest and most established styles and relates to buying securities that appear undervalued, while selling those that seem to be overvalued, with origins dating back to the first half of the 20th century (Graham, 1949). It is typically implemented by buying securities that rank higher relative to a fundamental anchor, while selling the ones that rank lower.

In the DM sovereign fixed income space, the measure of value is typically represented by the real yield, defined as the nominal yield of a sovereign bond less inflation expectations (Asness et al. 2015). In practice, either survey-based inflation expectations, or past realised inflation, or a combination of both are used (e.g. Fattouche (2018)). Since our sample is relatively large, we do not have the same availability and quality of inflation forecasts for all 18 countries throughout the full period. Given our preference for consistency, we construct our value measure as the 10 year bond yield (monthly average) minus the year-over-year realised CPI growth, lagged 3 months to ensure point-in-time data availability and to allow for first revisions (with any subsequent revisions typically expected to be minor).[5]

A high value score implies that the bond is undervalued relative to its fundamentals (high real yield, low real bond price). Hence, we go long the countries with high value scores and short the ones with low real yields. The strategy should profit if nominal yields revert back towards fundamental levels, i.e. rise in higher inflation countries and fall in lower inflation ones. This should happen because higher inflation leads to fixed income becoming less attractive as an asset class since it erodes the purchasing power of nominal payments. Demand for bonds in higher inflation countries should therefore fall, leading to lower bond prices and higher yields.

Carry

Carry captures the returns investors can realize if market conditions—the yield curve in our case—stay approximately the same.

We use the "simple carry" implementation, which corresponds to the "term spread" (cf. Brooks and Moskowitz (2017), Koijen et al. (2017)). Given our focus on the 7-10 year maturities, we define the carry factor as the monthly average of the 10-year bond yield minus the local short rate (cf. footnote 4).

The strategy benefits if conditions stay the same, as the bond will move closer to its maturity date and command a lower yield/higher price in the following period. A steeper yield curve translates to a higher carry return, all else equal. Hence, we go long the sovereigns with the steepest yield curves and short those with the flattest ones.

5 We tested an alternative definition of value where we subtract maturity matched inflation expectations instead of the realised CPI for a smaller subsample of countries and a shorter timeframe to confirm that results are comparable and that using lagged realised instead of expected inflation does not materially affect performance.

Momentum

Momentum is a strategy that capitalizes on current trends continuing into the future. It was originally documented in academic literature in the early 1990s (Jegadeesh and Titman (1993)). Several theories have been proposed to explain this phenomenon and why such strategies may be profitable, including the investor herding behaviour, short-term underreaction and/or medium-term overreaction, etc. (cf. Barberis et al. (1998), Daniel et. al (1998), Hong & Stein (1999)).

In this paper, we implement the "cross-sectional momentum" strategy, which capitalizes on a dynamic where best performing sovereigns continue outperforming, and those performing poorly would continue underperforming.

Instead of focusing on a specific window to define the "past trend", we rank countries based on the prior 3, 6, and 12-month excess, duration-adjusted, rolling cumulative returns, and create a composite equal-weighted momentum benchmark factor based on the three momentum strategies. **FIGURE 2** demonstrates that this approach allows us to diversify away from the idiosyncrasies specific to each of the individual lookback windows.

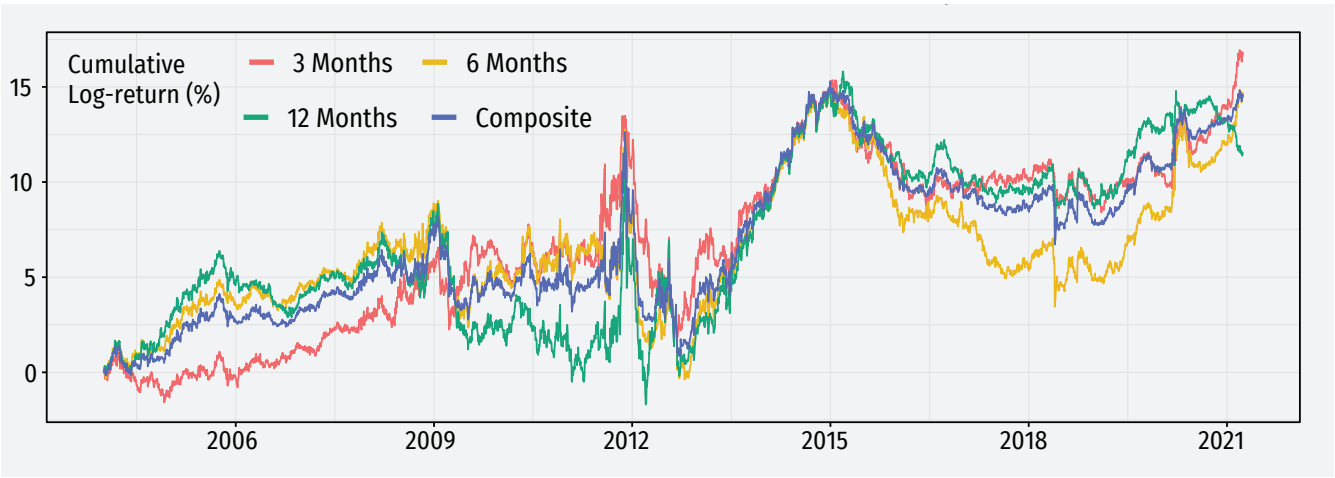


FIGURE 2: Momentum benchmarks, cumulative excess log-returns with different lookback windows
Sample: 18 countries, 7-10 year maturities, local currency, duration-scaled excess returns. *Source: RavenPack, June 2021*

Defensive

Defensive (or quality) style factor is meant to capture the tendency of safer assets to outperform lower quality, higher risk counterparts.

There is disagreement in the literature on how to define this style in the fixed income space. Some papers ignore it altogether and prefer to only focus on the value, carry and momentum styles (e.g. Fattouche (2018), Brooks and Moskowitz (2017)). Others define it as a pure maturity bet within each country, equal-weighted across countries (Brooks et al. 2018).

We prefer a cross-sectional definition of the defensive style, which enables us to directly test the incremental value-add of country-specific macroeconomic news sentiment in a way similar to the other three styles. Although we focus only on developed market sovereigns, **FIGURE 3** shows that there is substantial variability in creditworthiness within this cohort, especially considering that our universe includes countries like Italy and Spain, which experienced distress during the European Sovereign Debt crisis, as well as Japan with its extraordinarily elevated sovereign debt levels.

We implement the style by going long the highly rated sovereigns and shorting the low rated ones, investing equally in countries with the same ratings.

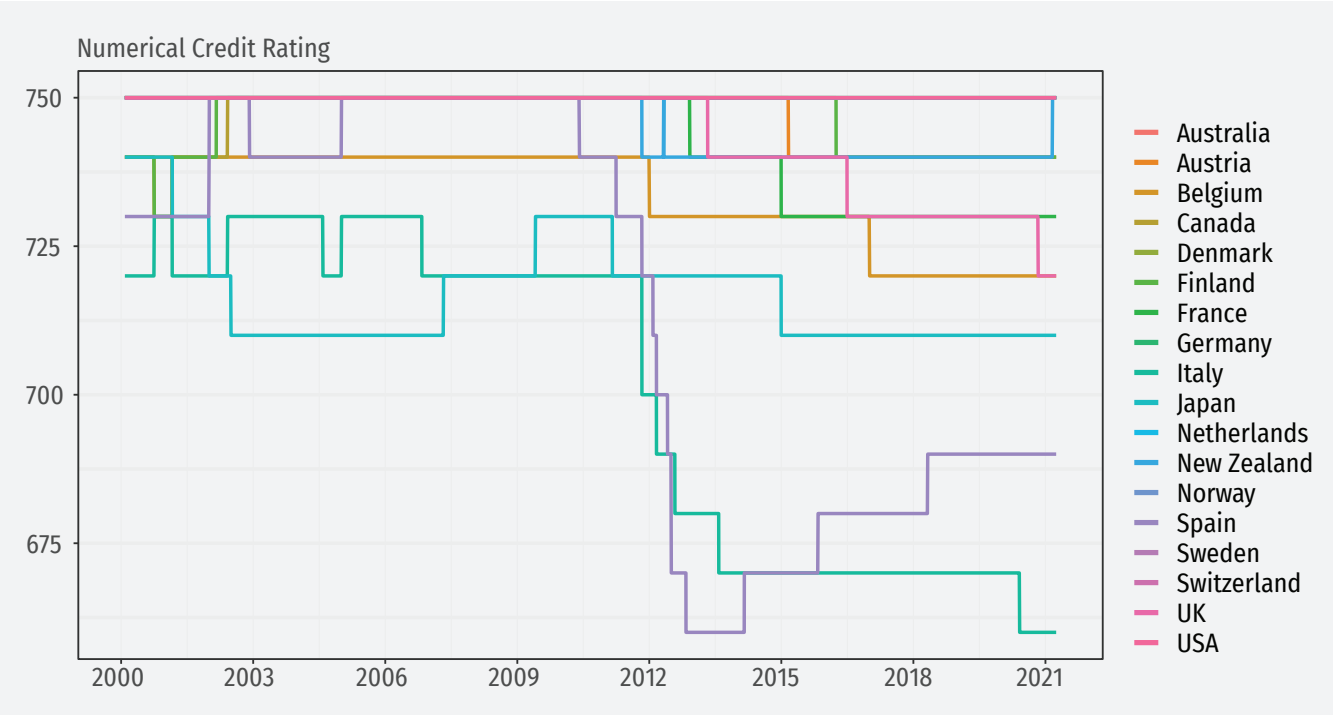


FIGURE 3: Numerical equivalents of ICE BofA composite ratings, 7-10 year maturities, government indices
Source: FactSet

2.3 Benchmark Construction & Performance

To construct the benchmark portfolios for each style, we proceed as follows:

- On the first trading day of each month, we rank countries based on the style scores available as of the last trading day of the previous month.
- We standardize the ranks by subtracting the cross-sectional mean rank and dividing by the standard deviation of ranks, where country i's weight in a style's benchmark portfolio at time t is:

$$W_{it} \propto \frac{Rank_{it} - avg(Rank_t)}{std(Rank_t)}, \; Rank_{it} = rank(Style_{it})$$

This transformation creates sets of positive and negative weights that sum to zero. It has been implemented in the literature as a more robust alternative to the standard "long x% – short x%" approach in an effort to mitigate the influence of outliers and improve diversification (e.g. Brooks and Moskowitz (2017), Asness et al. (2015)).[6]

We then scale each side of the strategy to one dollar long and one dollar short positions to ensure constant leverage. We also scale all returns to a constant duration of seven to ensure comparability across countries and over time, as well as duration-neutrality of the long-short style portfolios. We chose seven as it roughly approximates the duration of a ten-year DM government bond.

TABLE 1 summarizes the performance of our style benchmark portfolios over the January 2004 — March 2021 period, while **FIGURE 4** shows the cumulative excess log-returns.

	Sharpe Ratio	Ann. Return (%)	Ann. Volatility (%)	Max Drawdown (%)
Carry	0.37	1.34	3.63	12.89
Value	0.35	1.12	3.22	10.43
Defensive	-0.13	-0.51	3.93	27.70
Momentum	0.28	0.86	3.09	11.78
Combo	0.77	0.76	0.98	1.41

TABLE 1: Performance summary of style benchmark portfolios
Performance statistics based on the 02/01/2004 - 31/03/2021 sample, 18 countries, 7-10 year maturities. Local currency, duration-scaled excess returns used in all computations.
Source: RavenPack, June 2021

6 We also tested all strategies using the more standard long top tercile — short bottom tercile weights construction to confirm that all results still hold.

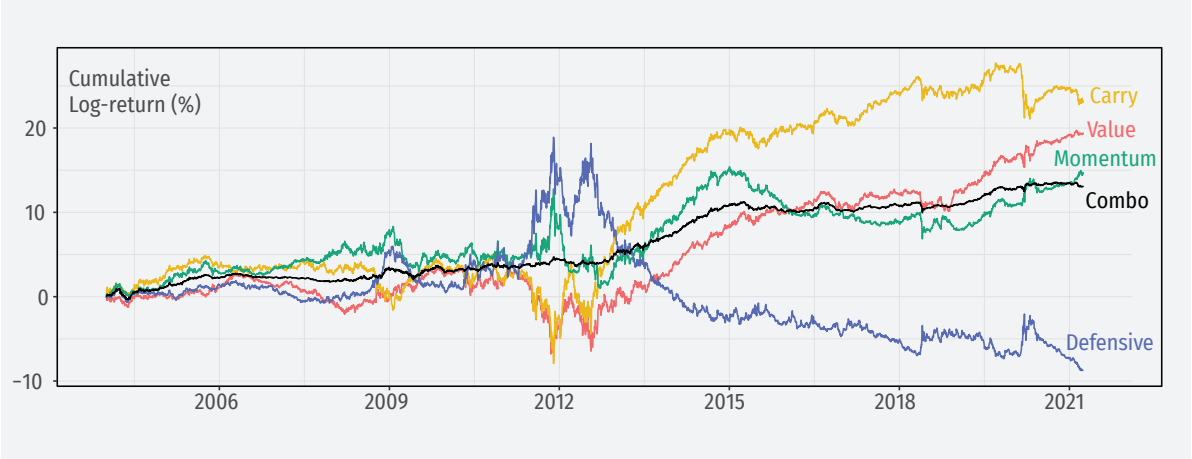


FIGURE 4: Cumulative excess log-returns of style benchmark portfolios
January 2004 – March 2021 sample, 18 countries, 7-10 year maturities. local currency, duration-scaled excess returns.
Source: RavenPack, June 2021

	Carry	Value	Momentum	Defensive
Carry	1.00	0.48	-0.27	-0.76
Value	0.48	1.00	-0.34	-0.74
Momentum	-0.27	-0.34	1.00	0.37
Defensive	-0.76	-0.74	0.37	1.00

TABLE 2: Benchmark style factor correlations
Correlations based on the 02/01/2004 – 31/03/2021 sample, 18 countries, 7-10 year maturities.
Local currency duration-scaled daily excess returns.
Source: RavenPack, June 2021

As expected, the carry factor exhibited significant drawdowns during periods of stress, such as the Great Recession, the 2010 — 2013 European Sovereign Debt crisis, the 2015 and 2018 China-related growth deceleration fears, as well as the most recent Covid episode, when market conditions changed drastically and capital flowed to low-yielding, but "safer" sovereigns.

These episodes precisely coincide with the outperformance of the defensive style factor. Indeed, the return correlations between carry and defensive styles is a significant −0.76 (cf. **TABLE 2**), which is to be expected, as the safest, highest rating sovereigns are precisely those that offer the least attractive yields, and therefore score poorly from a carry perspective. While the defensive style factor outperforms during periods of stress and offers substantial diversification benefits in the context of a combined style portfolio, its overall standalone performance has been lacklustre, with a negative Sharpe ratio over our sample period (**TABLE 1**). This is reasonable, since most

developed economies maintain high quality credits, hence DM sovereign fixed income is an unlikely space where the defensive factor can be expected to perform well on an ongoing basis.

The value factor performs similarly to carry, since inflation rates have remained under control, and generally exhibit high correlations across most developed economies throughout our sample period (cf. **FIGURE 5**). Hence, the value and carry style measures have, to a large extent, been driven by the behavior of 10 year yields.

FIGURE 4 also suggests that the momentum factor outperforms the value and carry factors during stress episodes, as prior performance trends persist into the future as the crises unfold. While momentum is somewhat negatively correlated to value and carry and positively correlated with the defensive factor, momentum performed well during the 2013–2014 episode when conditions normalised after the European sovereign debt crisis. Indeed, the performance of carry and value factors reverted around July 2012 after Draghi delivered his "whatever it takes" speech. Momentum followed starting in late 2012, as the style increased its exposure to the higher-yielding, less safe outperformers, while the defensive factor continued loading up on the low-yielding, higher quality sovereigns.

The combined equal-weighted strategy benefits meaningfully from the strongly negative correlations between value/carry and defensive styles, as well as from the low correlations between momentum and the other three styles. With the Sharpe ratio more than double the magnitude of any other standalone style and the max drawdown falling to as low as 1.4%, the combined strategy realizes significantly better risk-adjusted returns.

Evidently, combining the style factors results in improved performance due to significant portfolio diversification benefits. Next, we test whether introducing sentiment can further enhance risk-adjusted returns by allowing for a more dynamic style factor implementation.

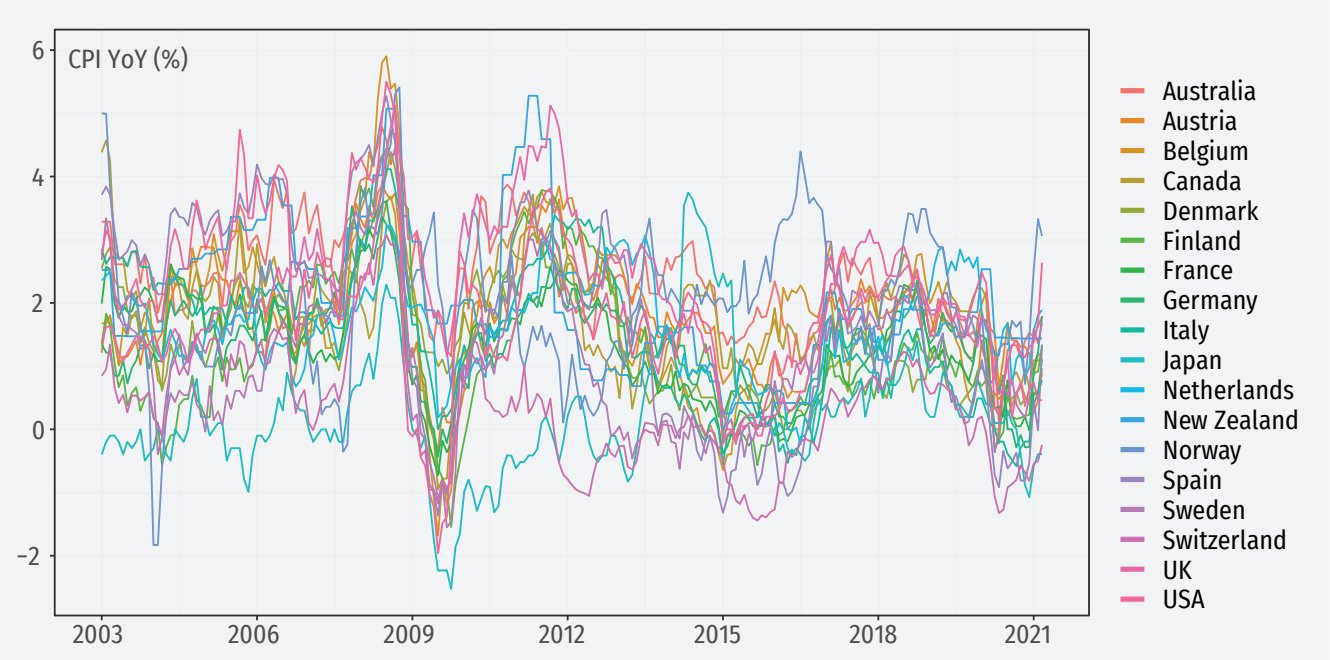


FIGURE 5: Inflation rates (YoY CPI, %)
Source: Macrobond

3. Sentiment-enhanced Style Factors

3.1 News Filtering

RavenPack has developed proprietary natural language processing (NLP) algorithms to identify almost 7,000 different event categories and over 300,000 entities in news articles. These algorithms are deployed daily in real time on thousands of news articles with archives going back to 2000.

The first step in building an appropriate news-based signal is to define a relevant news universe. The filtering we apply in this paper is similar to the one used in our recent FX style investing study (Hafez et al. (2020)). Given an analogous scope at macro and country level, our goal is to maintain consistency and minimize concerns of data snooping.

For each country in our dataset, we apply the following filters to select a subset of relevant news at the macro level:

1.	Topic	Economy
2.	Entity Type	Place or organization
3.	Relevance	≥ 90
4.	Event Relevance	≥ 90
5.	Event Similarity Days	≥ 30

The first two criteria ensure that we isolate macroeconomic news. Given that DM sovereign bonds are issued in their local currencies and are considered free of default risk, the primary driver of returns can be reasonably assigned to interest rate risk. Economy-related news can directly impact the outlook for interest rates. In order to eliminate any records associated with micro entities, such as companies, products or people, we only screen for events directly related to places and organisations.

Criteria 3, 4, and 5 aim to screen for the most relevant and novel news events. In particular, the 3rd and 4th criteria ensure that the event and the entity are identified in the headline of the news article, while the 5th criterion filters out duplicate entries, retaining only the first instance of the event identified within a period of 30 days, a period corresponding to the frequency of most economic data releases.[7]

FIGURE 6 shows the resulting relative occurrence of news aggregated at the group level by country. As we can see, the most important categories are consumption, employment and production. The distributions seem relatively stable across regions, with a logical mix of events within each country.

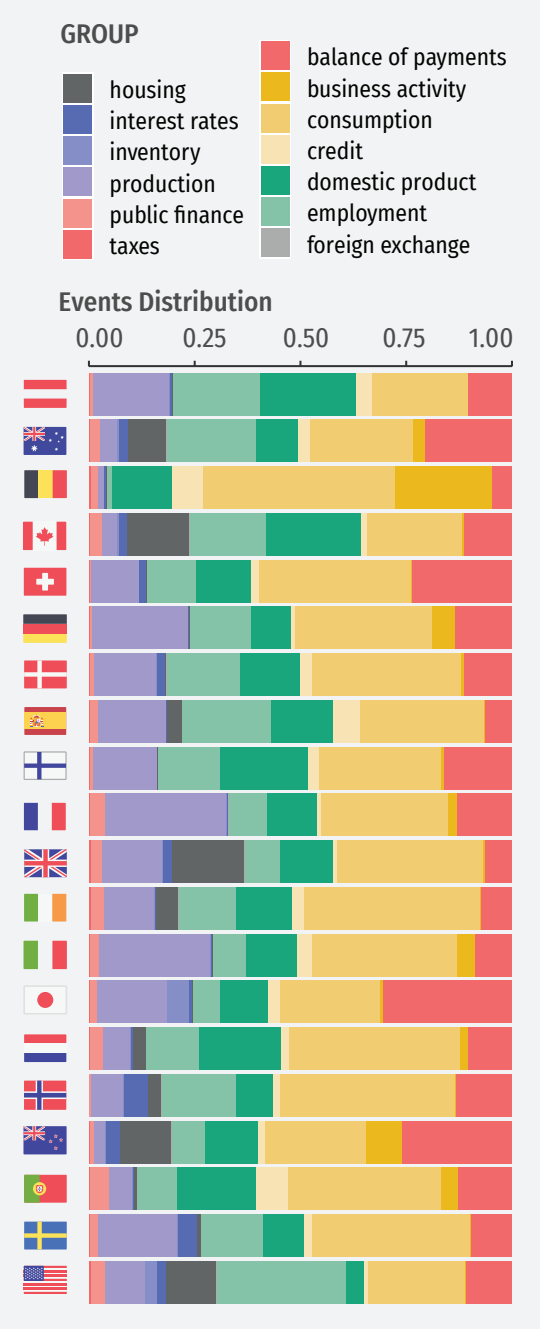


FIGURE 6: News distribution by country aggregated at the events group level
Source: RavenPack, June 2021

7 Results are robust to using a quarterly deduplication window (ESD >=90) as in our FX paper (Hafez et al. 2020).

3.2 Sentiment Signal Construction

For each event identified, RavenPack produces a wide range of analytics, including several sentiment scores. In this paper, we focus on the Composite Sentiment Score (CSS).[8] CSS is a document-level score that has been trained by applying several NLP and expert consensus techniques to equity market intraday data to determine the sentiment modality of a news story.[9]

News around improvements in economic data, the expectations of easier financial conditions, fiscal and monetary stimulus, all have a tendency to elevate equity market sentiment and hence command positive CSS scores. For applications in the fixed income space, the CSS needs to be inverted, as bonds in countries where growth prospects are better and discount rates are expected to be lower (higher CSS) should be relatively less attractive than in countries where the reverse is the case (lower CSS).

Sentiment metrics tend to exhibit high volatility over daily frequencies. Using daily scores could be helpful in the case of fast moving, highly liquid markets such as FX. In the fixed income space, however, capital tends to be deployed at a slower pace, resulting in a more gradual rate at which country-specific information is incorporated into sovereign bond prices. We therefore refrain from using the daily aggregation window when defining our sentiment metrics and focus instead on applying the smoothing windows of 3, 5, 10, and 21 days.[10]

For each country i and for a given smoothing window of N days ($N \in \{3, 5, 10, 21\}$), we calculate a daily sentiment indicator as follows:

1. Create a daily rolling volume V weighted moving average (MA) of the inverted CSS:

$$\frac{\sum_{t=j-N}^{j} -CSS_t \times V_t}{\sum_{t=j-N}^{j} V_t}$$

2. Create the standardized sentiment score S_{it} by computing a daily rolling z-score of this MA sentiment signal, using the trailing 3-year window for the mean and standard deviation estimates.

In addition to the four strategies with different smoothing windows, we also consider a hybrid version (Multi-CSS), which equally invests across each of these four strategies. In essence, the Multi-CSS strategy introduces a time decay because more recent days receive higher weights as they appear in all smoothing windows, while older sentiment only affects longer aggregation windows.

The New Zealand market is the first in our sample to close on day t, so we define a day as the 24h preceding 4pm Pacific time. This ensures that no lookahead bias is introduced when using the closing prices in our analysis. Given that the local market closes at 5pm Pacific time, we leave one hour of margin for trading.[11]

8 Similar to our FX paper (Hafez et al. 2020), we find that CSS-based tilts have a more robust, stable, and better performance than ESS-based overlays.

9 CSS combines five sentiment scores using an intuitive set of rules that ensures no sentiment disagreement exists among them. All training and testing was conducted on randomly selected news articles and concluded by July 2007.

10 To confirm our intuition, we also implemented the 1-day based overlays and found that they lead to worse standalone performance. However, including them within the multi-sentiment construct does not hurt performance substantially.

11 As a robustness check, we also tested a 5pm New York cut-off time and trading using next day's closing prices, with similar results.

3.3 Sentiment Overlays

We overlay the sentiment signals S_{it} on the benchmark factor weights constructed in Section 2.3. as follows:

- Given the long-side weights of a factor benchmark $W_{it} > 0$, we define the new sentiment-tilted weights W^*_{it} as:

$$W^0_{it} = W_{it} \exp(aS_{it}),\ W^*_{it} = \frac{W^0_{it}}{\sum_{i,W^0_{it}>0} W^0_{it}}$$

  where a is a scalar constant used as a scaling factor to control the tilt magnitude.

- Similarly, we construct the short-side tilted weights:

$$W^0_{it} = W_{it} \exp(-aS_{it}),\ W^*_{it} = \frac{W^0_{it}}{\sum_{i,W^0_{it}<0} |W^0_{it}|}$$

While sentiment is available daily, certain asset managers may not be able or willing to rebalance at such high frequency. Therefore, we also test whether sentiment tilt performance enhancements persist at weekly and monthly rebalancing.[12]

For illustration, **FIGURE 7** shows the tilts in the weights of the U.S. carry style factor with monthly rebalancing at different scaling values. The blue line corresponds to the benchmark weight with no sentiment overlay. As expected, we observe the following:

- sentiment only changes the magnitude of the weights without affecting the decision of whether to go long or short the sovereign
- larger scaling values lead to increased sentiment tilts and moving further away from the benchmark

FIGURE 14 in the Appendix compares daily, weekly, and monthly rebalancing weights when scaler a is set to one.

While higher frequency rebalancing and larger sentiment scalers clearly generate more portfolio turnover, they also allow us to implement the style factors in a more dynamic way. The question we address in the following section is whether this more dynamic implementation enhances or detracts from performance.

While we set the scaler to 1 when showcasing the main results, **FIGURE 19** in the Appendix shows that results remain robust when using higher or lower sentiment scaler values.

12 Monthly sentiment-based rebalancing entails using the sentiment tilted weights instead of the default benchmark weights on the rebalance day. For the weekly rebalancing strategy, we chose to rebalance every Wednesday, but results remain robust to trading on a different day.

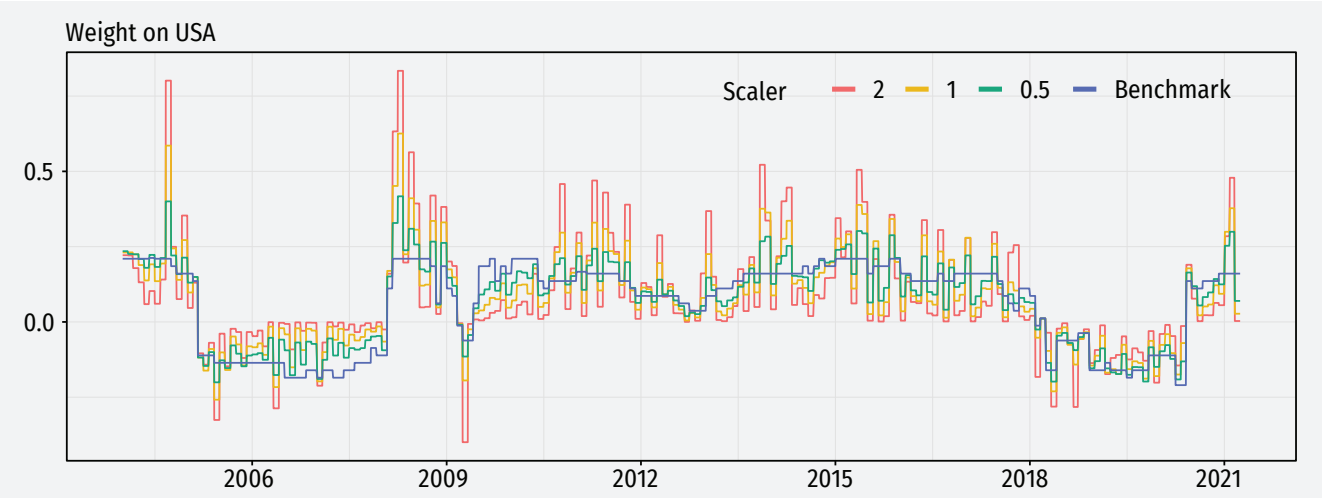


FIGURE 7: U.S. Carry style weights Benchmark with no sentiment (blue) versus the benchmark with the sentiment (multi-CSS) overlay, using monthly rebalancing frequency across different scaler values. *Source: RavenPack, June 2021*

4. Results

4.1 Main Sample

Our main sample corresponds to sovereign bonds with 7-10 year maturity, across 18 countries, and using local currency excess returns.

We find that sentiment tilts enhance performance across most sentiment aggregation windows, and therefore decided to focus on the Multi-CSS overlays for simplicity and clarity (see Section 3.2 for details on sentiment signal construction and **FIGURES 15** and **16** in the Appendix for a more granular, disaggregated performance overview).

FIGURE 8 contrasts the cumulative excess log-returns of the Multi-CSS overlay strategies across the four different style factors, as well as for the equal-weighted benchmark, while **FIGURE 9** shows the corresponding Sharpe ratios.

Although sentiment overlays evidently perform better with daily and weekly rebalancing, monthly tilts also enhance risk-adjusted returns for each standalone style factor.

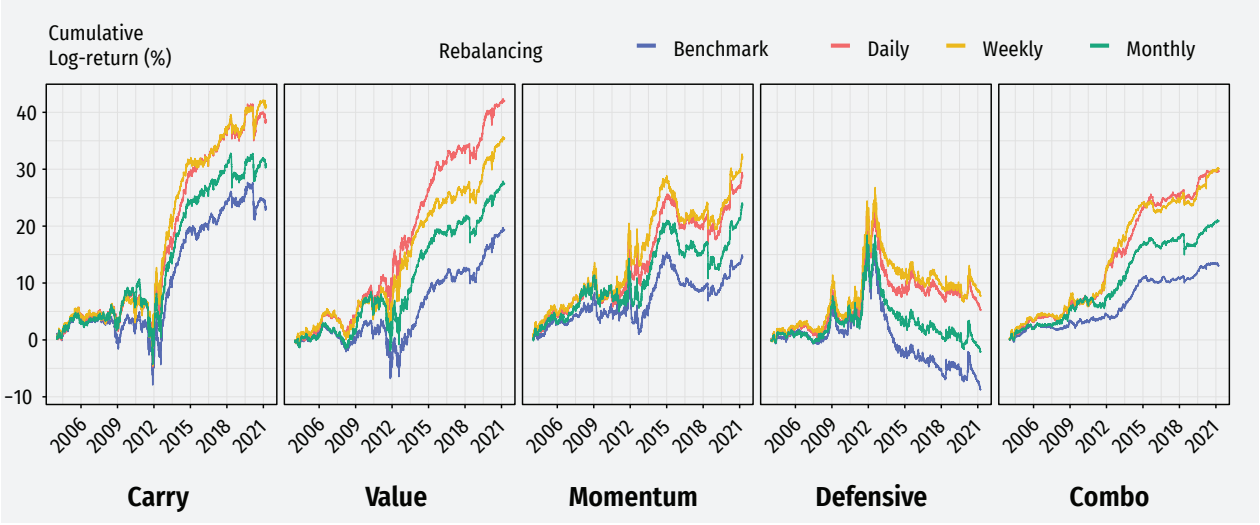


FIGURE 8: Cumulative excess log-returns of sentiment-enhanced styles
Main results: 18 countries, 7-10 year maturities, local currency, based on multi-CSS overlays, disaggregated by style, and rebalancing frequency.
Source: RavenPack, June 2021

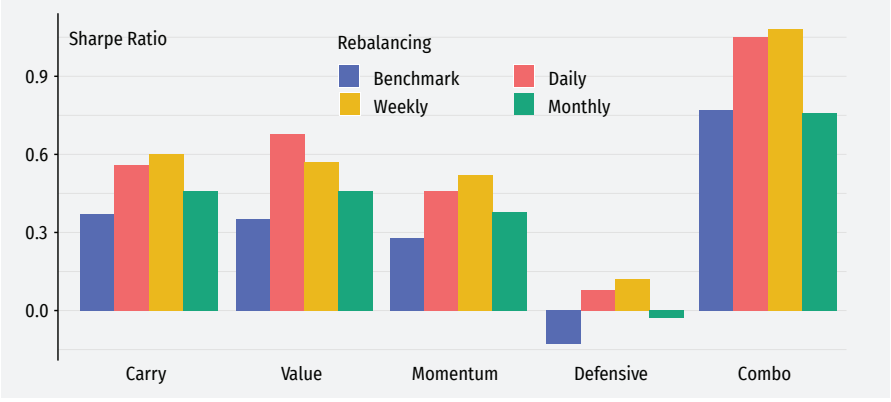


FIGURE 9: Sharpe ratios of sentiment-enhanced styles (multi-CSS overlays)
Main results: 18 countries, 7-10 year maturities, local currency, based on multi-CSS overlays, disaggregated by style, and rebalancing frequency.
Source: RavenPack, June 2021

TABLE 3 shows that, except for carry, daily and weekly sentiment tilts also significantly reduce maximum drawdowns across all standalone styles.

Applying sentiment tilts to the defensive style reduces maximum drawdown by 8 percentage points and improves performance, increasing the Sharpe ratio from –0.13 to +0.08 and +0.12 when rebalancing daily and weekly, respectively.

For the value factor, daily sentiment-based rebalancing almost doubles the Sharpe ratio from 0.35 to 0.68, while reducing the maximum drawdown by almost a half.

Carry and momentum factors perform best with weekly rebalancing, boosting the Sharpe ratios by approx. 60% and 80%, respectively.

	Rebalancing	Sharpe Ratio	Annual Return (%)	Annual Volatility (%)	Max Drawdown (%)
Carry	Benchmark	0.37	1.34	3.63	12.89
	Daily	0.56	2.23	3.96	13.22
	Weekly	0.60	2.38	3.99	12.36
	Monthly	0.46	1.77	3.87	14.94
Value	Benchmark	0.35	1.12	3.22	10.43
	Daily	0.68	2.43	3.57	6.54
	Weekly	0.57	2.05	3.63	8.90
	Monthly	0.46	1.59	3.44	10.43
Defensive	Benchmark	-0.13	-0.51	3.93	27.70
	Daily	0.08	0.30	3.80	19.06
	Weekly	0.12	0.45	3.76	20.12
	Monthly	-0.03	-0.12	3.69	20.58
Momentum	Benchmark	0.28	0.86	3.09	11.78
	Daily	0.46	1.69	3.63	9.14
	Weekly	0.52	1.87	3.60	9.46
	Monthly	0.38	1.39	3.66	10.25
Combo	Benchmark	0.77	0.76	0.98	1.41
	Daily	1.05	1.72	1.63	2.25
	Weekly	1.08	1.75	1.62	2.43
	Monthly	0.76	1.21	1.59	3.73

TABLE 3: Sentiment-enhanced styles (multi-CSS overlays) performance vs benchmarks
Performance statistics based on the 02/01/2004 - 31/03/2021 sample, 18 countries, 7-10 year maturities. Local currency duration-scaled excess returns used in all computations.
Source: RavenPack, June 2021

The combined strategy, which in its basic form has more than double the Sharpe ratio of any standalone style, sees further performance enhancements, increasing the Sharpe ratio from 0.77 to 1.05 and 1.08 when applying daily and weekly sentiment tilts, respectively. While the monthly rebalancing strategy improves annualised returns from 76 bps to 121 bps, it comes at the cost of an increase in volatility, leaving the Sharpe ratio unchanged.

The increases in volatility and maximum drawdown observed in the combined strategy are reasonable, given that the same sentiment signal is used across all styles. Incorporating this signal introduces further correlation to the styles relative to the well-diversified benchmark. However, both annualized volatility and maximum drawdown metrics for the combo strategy remain fairly low in absolute terms across all rebalancing frequencies.

In fact, sentiment tilts lead to significant increases in annualised returns without substantially raising volatilities. This is, in and of itself, an important advantage in the fixed income space, where significant leverage often has to be deployed to achieve a certain return hurdle, and is especially meaningful for investors who work under strict leverage constraints.

In summary, we find that there is substantial value in implementing style factors more dynamically by introducing sentiment tilts. **FIGURE 10** demonstrates how this dynamism improves performance without introducing noise. We split the portfolio into two groups based on the direction of the multi-CSS sentiment tilts and assign weights to each sovereign bond based on its overweight/underweight magnitude relative to the benchmark. We observe that both groups exhibit positive performance, which implies that we overweight countries precisely when they outperform their peers, and we also underweight them when they are lagging in their peer group.

The overweight portfolio exhibits consistent performance over time, while the underweight portfolio performs especially well during the Eurozone Sovereign Debt crisis. This indicates that sentiment would have provided significant portfolio protection during this crisis period by increasing the short positions in distressed sovereigns relative to the benchmark signal.

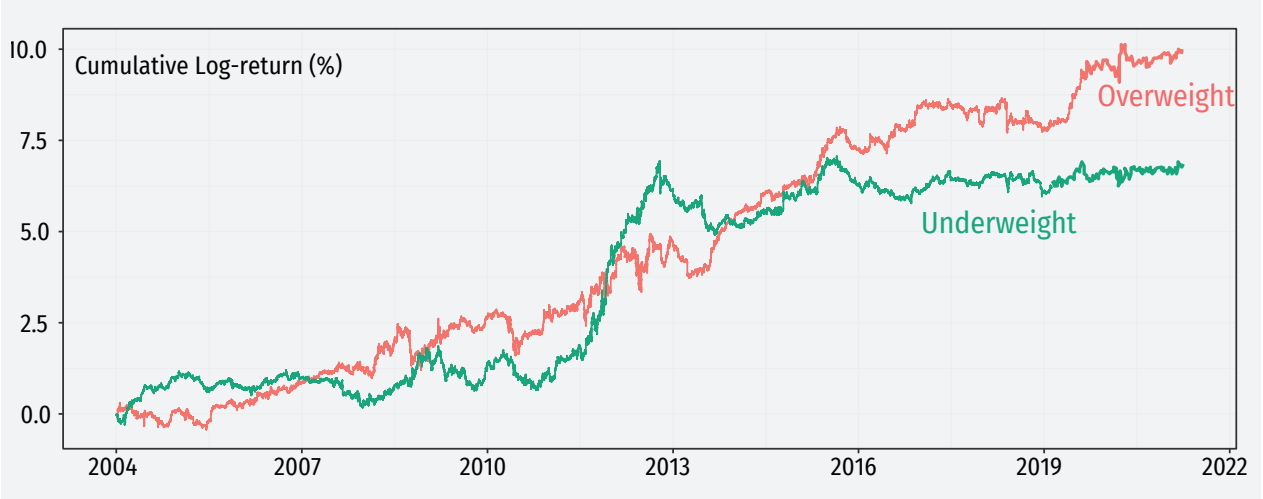


FIGURE 10: Cumulative excess log-returns of multi-CSS Combo strategy broken down by over/underweight tilts *Source: RavenPack, June 2021*

FIGURE 11 shows the relative proportions of weight tilts applied to the original benchmark weights by country over the entire backtesting period. It offers a confirmation that the outperformance is robust, with fairly even tilt distributions across the countries. Namely, outperformance cannot be attributed to luck in timing the exposures to a few sovereigns.

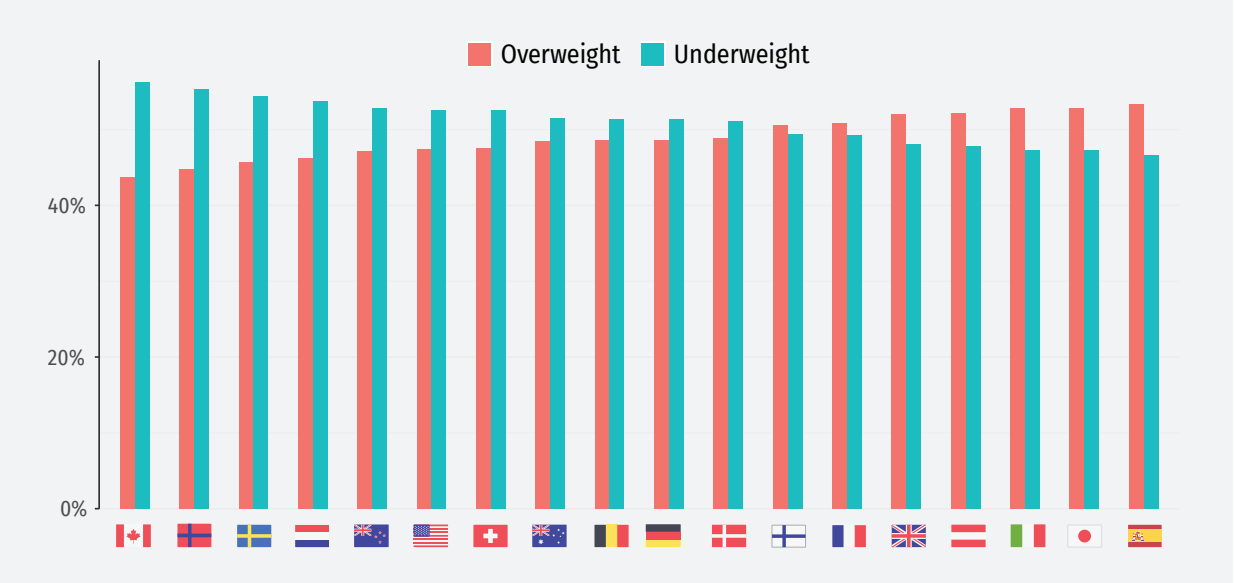


FIGURE 11: Over/Underweight split by country in the multi-CSS Combined styles ("Combo") strategy
Source: RavenPack, June 2021

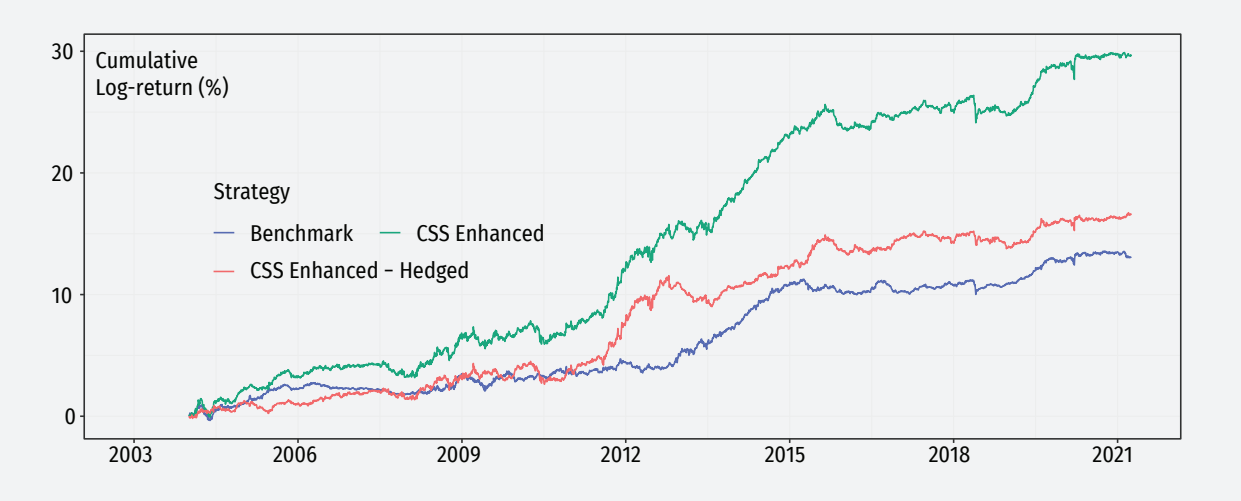


FIGURE 12: Cumulative excess log-returns of the multi-CSS enhanced Combo strategy hedged by shorting the combined styles benchmark *Source: RavenPack, June 2021*

Finally, **FIGURE 12** shows that the outperformance of the sentiment-enhanced strategy is consistent over time. In particular, hedging the Multi-CSS Combo strategy by shorting the corresponding benchmark delivers an annualised return of 95 bps with a 0.74 Sharpe ratio. **FIGURES 17** and **18** in the Appendix confirm that all key results are robust when using USD hedged excess returns or all-maturity indices instead.

4.2 Other Country Universes

As mentioned in Section 2.1, the literature generally focuses on specific subsets of 6 and 13 countries, depending on whether 10-year bond futures or the J.P. Morgan Government Bond Indices are used to test the style strategies. We test whether our results still hold when implemented on these subsamples. Additionally, we test our results on the subsample of European sovereigns.

FIGURE 13 shows that results remain attractive across all three subsamples. In particular, sentiment seems especially valuable in the 6-country subset, which resembles the most liquid investment universe and is therefore the most likely candidate for practical implementation. Within this highly liquid subset, sentiment overlays enhance performance across all rebalancing frequencies, for all standalone styles, as well as for the combined strategy. In the latter case, the Sharpe ratio almost doubles when implementing styles more dynamically with daily/weekly rebalancing frequencies. Even when rebalancing on a monthly basis, the Sharpe ratio increases from 0.43 to 0.58 in local currency terms.

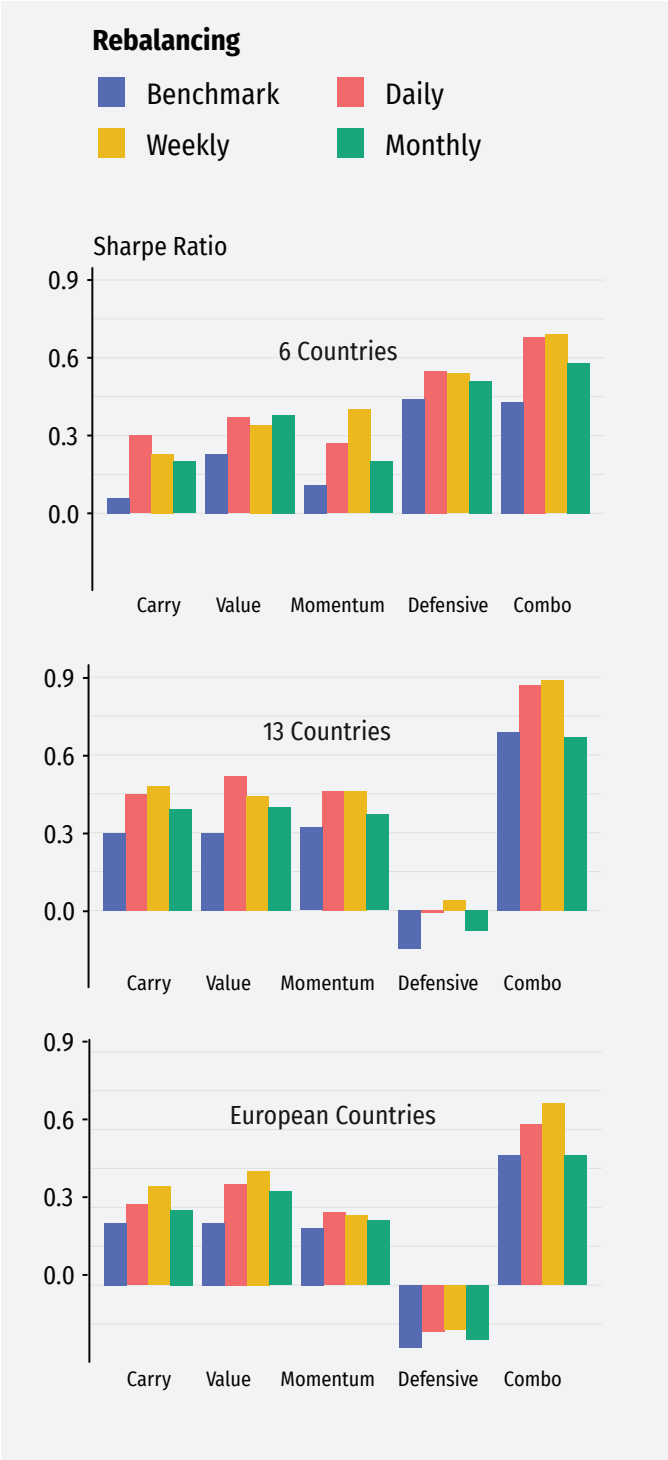


FIGURE 13: Sentiment-enhanced style strategies (multi-CSS overlays) evaluated on different country subsamples

Multi-CSS enhanced style factors vs. the benchmark, 7-10 year maturities.

Country subsamples:
- 6 countries (most liquid bond futures): Australia, Canada, Germany, Japan, UK, USA
- 13 countries: above 6 + Belgium, Denmark, France, Italy, Netherlands, Spain, Sweden
- European Sovereigns: Austria, Belgium, Denmark, France, Germany, Italy, Netherlands, Spain, Sweden, UK, Finland

Source: RavenPack, June 2021

5. Conclusion

While alternative data sparks excitement among finance practitioners, many asset managers have been struggling to find ways of incorporating it into existing investment frameworks.

In this paper, we take a key asset class — developed-market sovereign fixed income, and one of the most popular investing approaches — style investing, and show how news sentiment can be seamlessly introduced to enhance strategy performance.

We find that economic news sentiment improves risk-adjusted returns of standard value, carry, momentum, and defensive factors at different rebalancing frequencies and across different country universes. In particular, for the sample of 18 developed-market sovereigns, weekly sentiment-based tilts to default benchmark weights enhance the Sharpe ratio of the multi-style portfolio from 0.77 to 1.08 and more than double the annualised excess return from 76 bps to 175 bps.

While this paper focuses only on the Developed Markets, our next step is to apply a similar approach to Emerging Market bonds where, in addition to the interest rate risk, credit risk also comes into play. Indeed, political- and ESG-based news sentiment could play a significant role in addition to economic events.

Having already successfully tested a similar approach to enhancing the performance of style factors in the FX space, we also plan to investigate whether sentiment overlays can add value to commodities and global equity styles using this methodology.

6. Appendix

6.1 Weights at different rebalancing frequencies

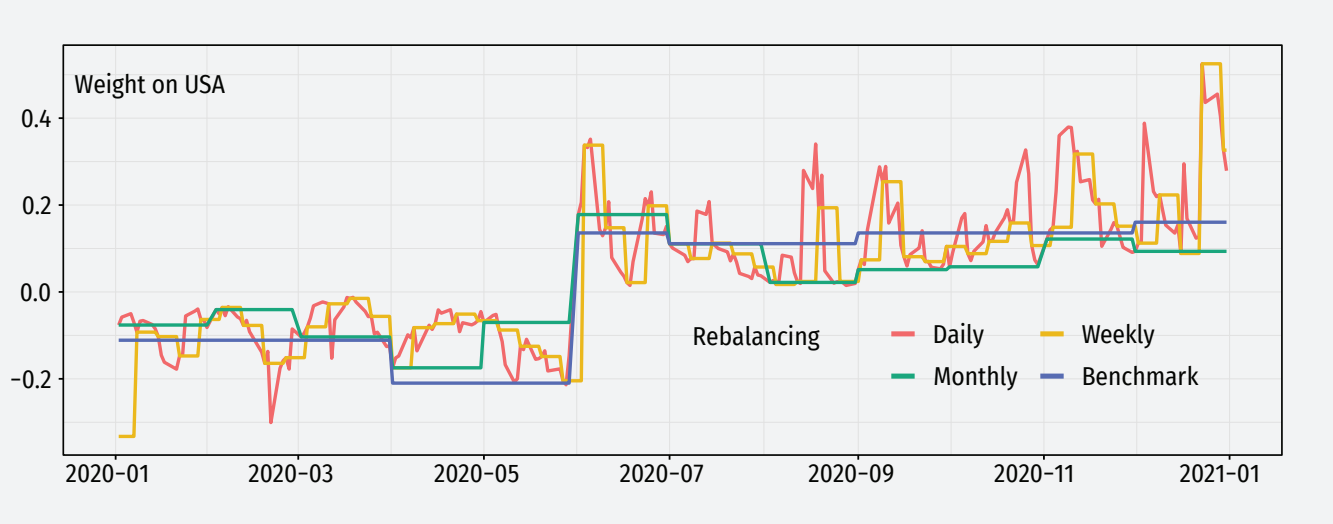


FIGURE 14: U.S. Carry style weights at different rebalancing frequencies
Benchmark with no sentiment (blue) versus the benchmark with the sentiment overlay, using monthly, weekly, and daily rebalancing frequencies. *Source: RavenPack, June 2021*

6.2 Performance disaggregated by sentiment window smoothing period

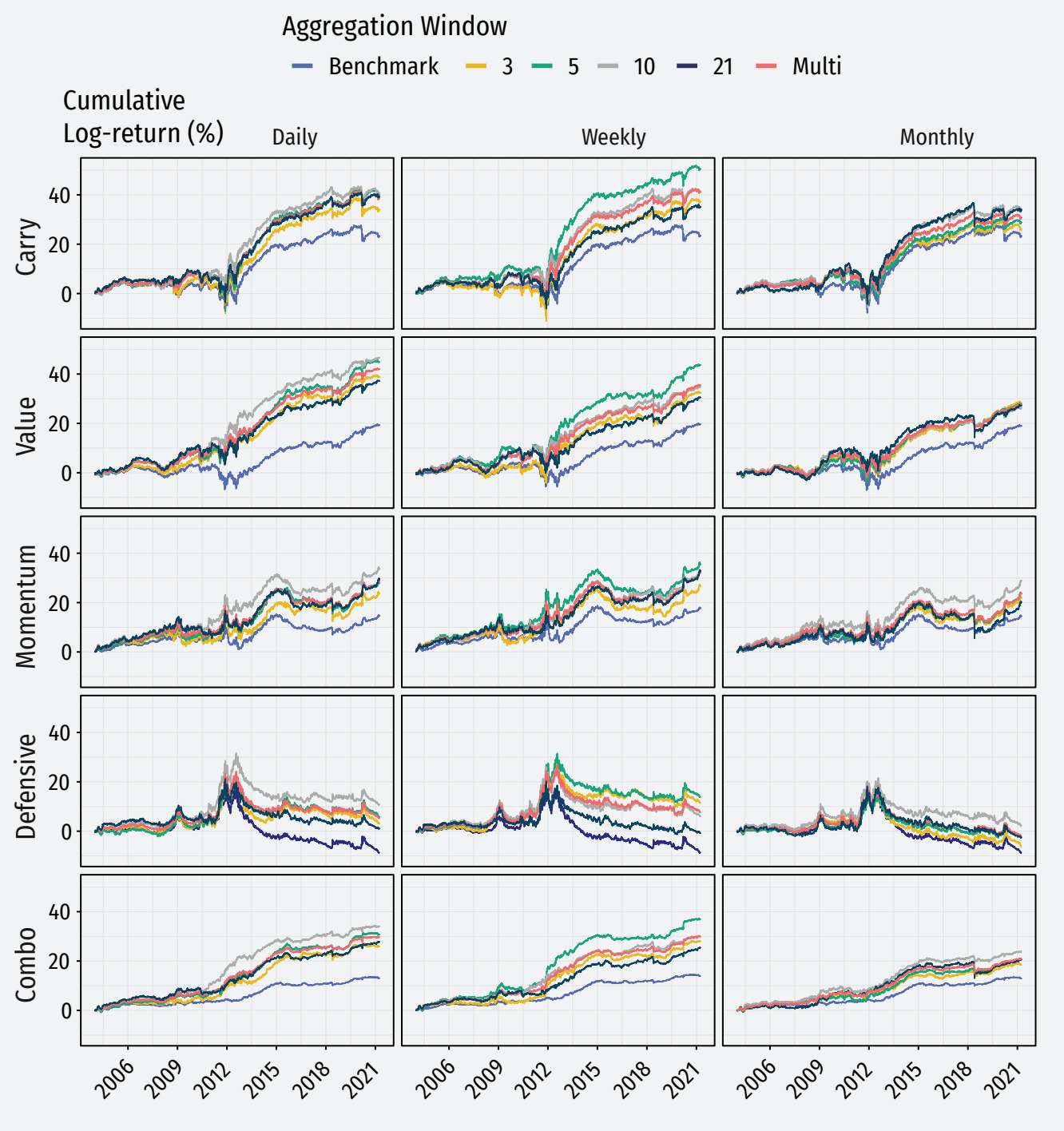


FIGURE 15: Cumulative excess log-returns of sentiment-enhanced styles
Main results: 18 countries, 7-10 year maturities, local currency, disaggregated by style, rebalancing frequency, and sentiment aggregation window. *Source: RavenPack, June 2021*

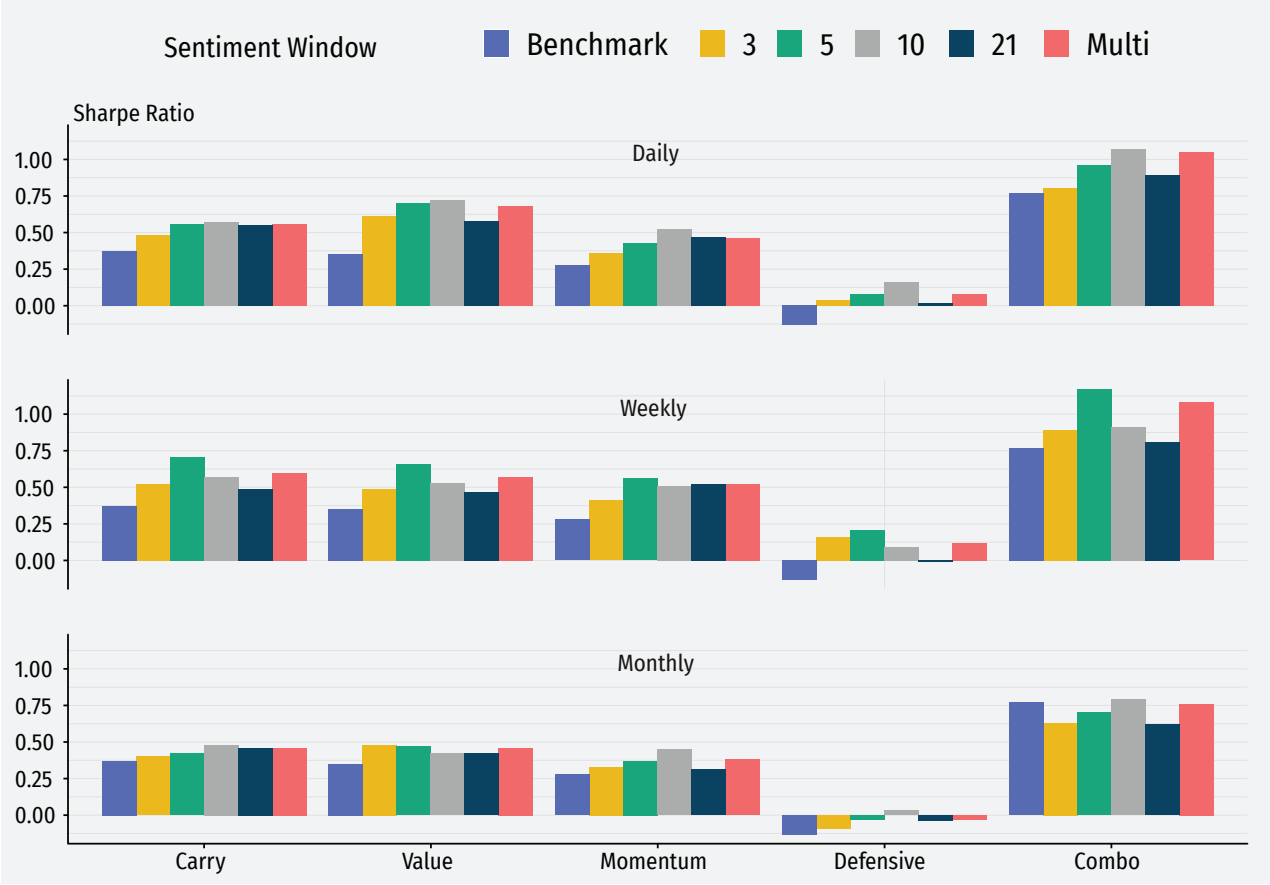


FIGURE 16: Sharpe ratios of sentiment-enhanced styles
Main results: 18 countries, 7-10 year maturities, local currency, disaggregated by style, rebalancing frequency, and sentiment aggregation window. *Source: RavenPack, June 2021*

6.3 USD Hedged Returns

The following figures use USD hedged excess returns instead of the local currency excess returns. In practice, portfolio managers may decide to hedge or not, or hedge partially or tactically depending on their FX views. In any case, all results still hold with similar outperformance for sentiment enhanced styles.

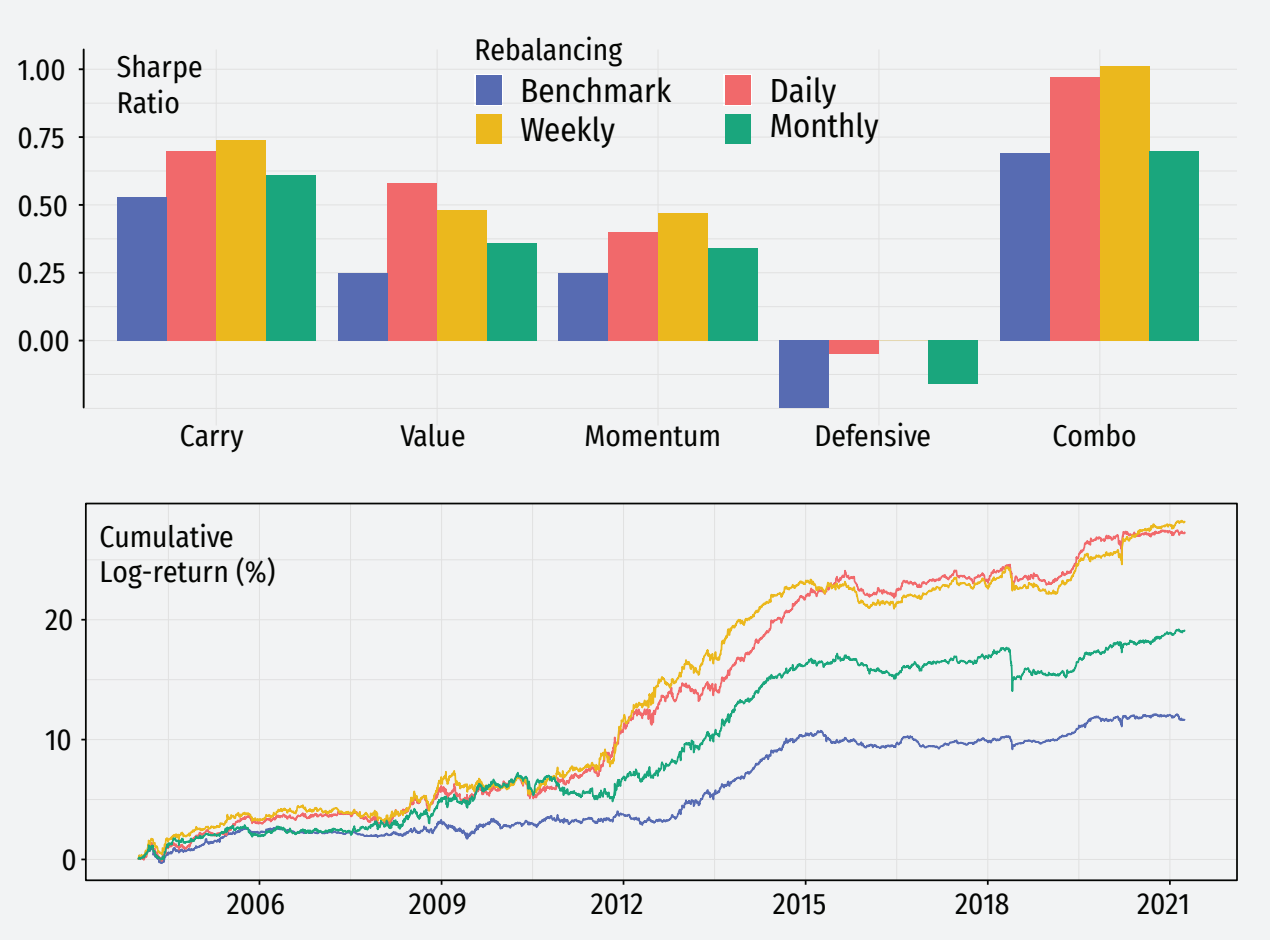


FIGURE 17: Sharpe ratios of sentiment-enhanced styles (multi-CSS overlays) and cumulative excess log-returns of the combined styles strategy. USD Hedged excess returns.
18 countries. 7-10 year maturities.
Source: RavenPack, June 2021

6.4 All Maturities

In addition to the 7-10 year bond indices, we test our results on the "All Maturities" indices.

The results are largely the same. For the combined strategy, daily and weekly rebalancing increase the Sharpe ratio from 0.88 to 1.12 and 1.13 in local terms and from 0.80 to 1.05 and 1.07 in USD hedged terms, respectively. As already highlighted, monthly sentiment-based rebalancing underperforms the benchmark for the combined strategy, but still adds value for each style on a standalone basis.

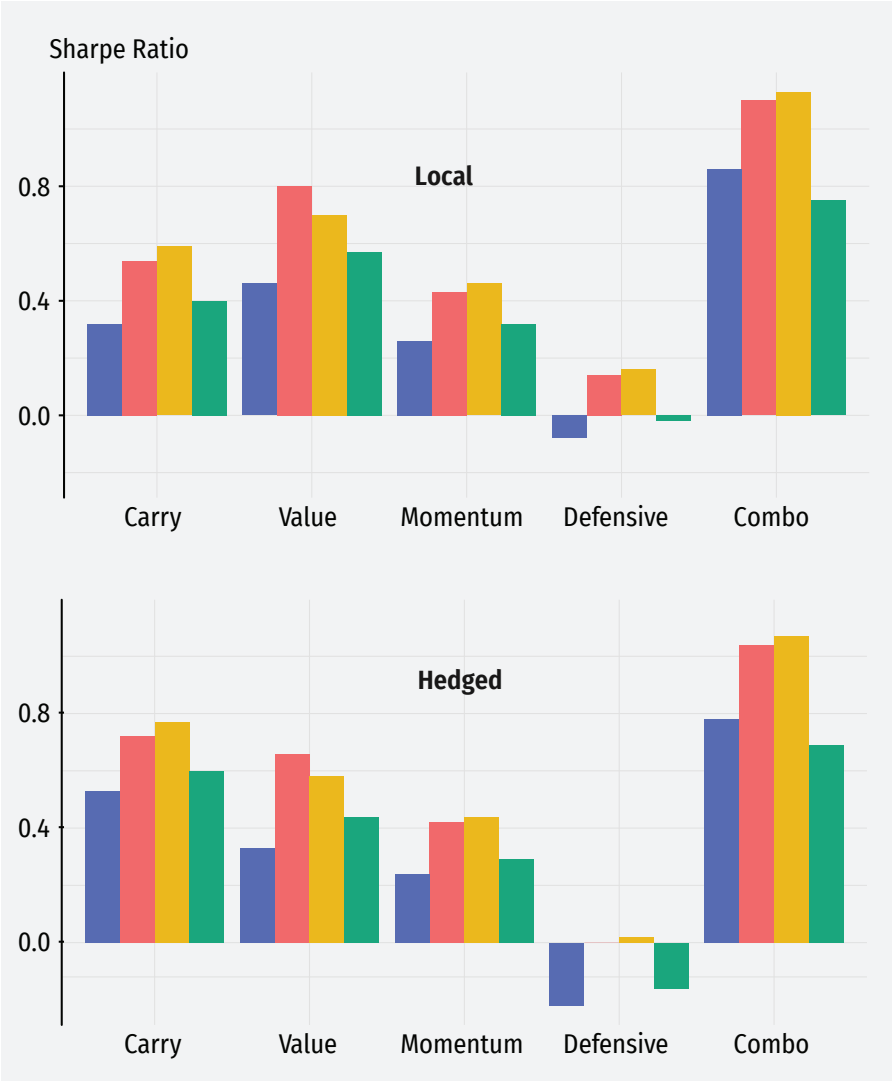


FIGURE 18: Sharpe ratios of sentiment-enhanced styles (multi-CSS overlays). All Maturities ICE BofA Government Bond Indices.
Source: RavenPack, June 2021

6.5 Different Scalers

We used a unit scaler when implementing sentiment tilts in all of the sentiment-based strategies presented in the paper. However, we can leverage sentiment to a lesser or greater extent by adjusting the scaler higher or lower.

FIGURE 19 shows that a higher scaler value improves performance for certain standalone styles, at certain rebalance frequencies (e.g. raising the scaler to 2 turns around the performance of the Defensive style at the monthly rebalancing frequency). However, the combined strategy performance remains largely unchanged, especially at daily and weekly rebalance frequencies.

Investors, conscious of transaction costs, may therefore still be able to take advantage of sentiment overlays within their style investing strategies, even if tilting the weights at a relatively small scale.

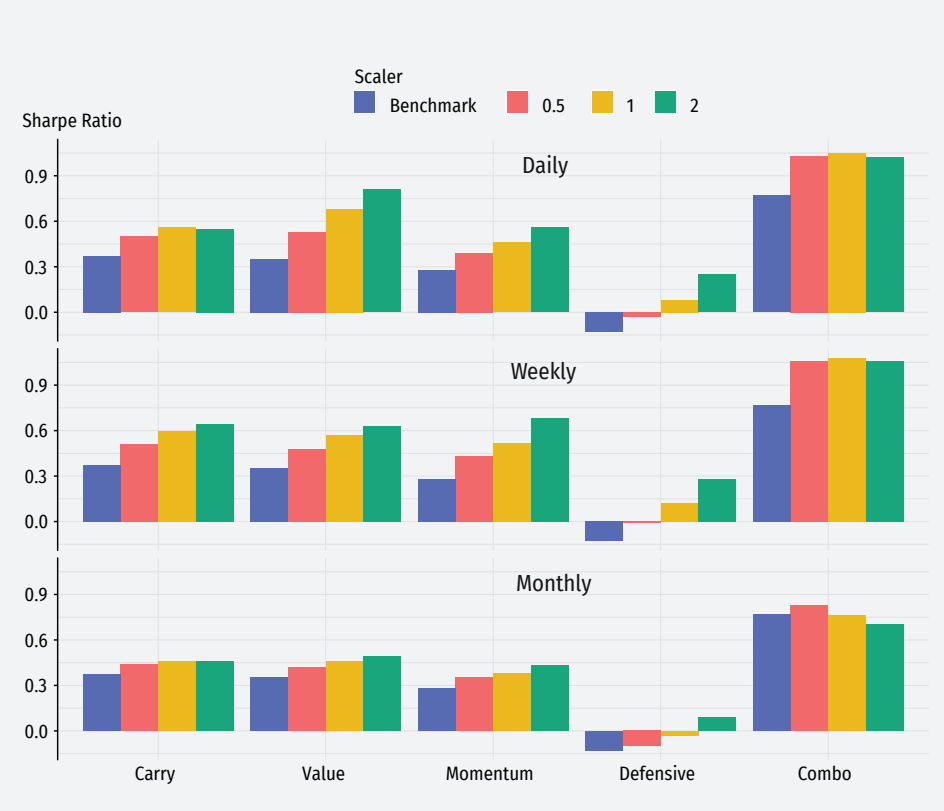


FIGURE 19: Sharpe ratios of sentiment-enhanced styles (multi-CSS overlays) at different rebalancing frequencies for different scaler values.
7-10 year maturities. Local currency returns. 18 countries sample.
Source: RavenPack, June 2021

References

Asness, C., Ilmanen, A., Israel, R., and Moskowitz, T. J. (2015). "Investing with Style." *Journal of Investment Management* 13 (1): 27–63.

Barberis, N., Shleifer, A., and Vishny, R. (1998). "A Model of Investor Sentiment." *Journal of Financial Economics,* 49(3), 1 September, 307–343.

Brooks, J., and Moskowitz, T. J. (2017). "Yield Curve Premia."

Brooks, J., Palhares, D., and Richardson, S. H. (2018). "Style Investing in Fixed Income." *The Journal of Portfolio Management* 44 (4): 127–139.

Cai, Z. and Erlwein-Sayer, C. (2017). "Enhanced Corporate Bond Yield Modelling incorporating Macroeconomic News Sentiment", **http://dx.doi.org/10.2139/ssrn.3405977**

Cheng, Stephanie F. (2021). "The Information Externality of Corporate Financial Information in the Secondary State-Bond Market". *Forthcoming, Journal of Accounting Research*, **http://dx.doi.org/10.2139/ssrn.3321344**

Daniel, K., Hirshleifer, D., and Subrahmanyam, A. (1998). "A Theory of Overconfidence, Self-Attribution, and Security Market Under and Over-Reactions," *Journal of Finance*, 53, 1839–1885.

Erlwein-Sayer, C. (2018). "Macroeconomic News Sentiment: Enhanced Risk Assessment For Sovereign Bonds". *Risks*, 2018, 6 (4), 141; **https://doi.org/10.3390/risks6040141**

Fama, E. and French, K. (1992). "The Cross-Section of Expected Stock Returns," *Journal of Finance* 47(2), 427–465.

Fattouche, C. (2018). "Style Investing in Rates Markets." *Barclays QIS Insights*, February 2018.

Frazzini, A. and Pedersen, L. (2014). "Betting Against Beta," *Journal of Financial Economics* 111(1), 1–25.

Graham, Benjamin. (1949), *The Intelligent Investor*

Gava, J., Lefebvre, W., and Turc, J. (2020). "Beyond Carry and Momentum in Government Bonds", *The Journal of Fixed Income*, Spring 2020, 29 (4), 48-74; **https://doi.org/10.3905/jfi.2020.1.085**

Hafez, P., Liu, A. and Matas Navarro, R., "Forex Carry, Value, & Momentum Factors Boosted by Sentiment" (September 30, 2020). **https://app.ravenpack.com/research/forex-carry-value-momentum-boosted-by-sentiment/**

Hong, H. and Stein, J. (1999). "A Unified Theory of Under-reaction, Momentum Trading, and Overreaction in Asset Markets," *Journal of Finance*, 54, 2143–2184.

Jegadeesh, N. and Titman, S. (1993). "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency," *Journal of Finance* 48(1), 65–91.

Koijen, Ralph, Tobias J. Moskowitz, Lasse Pedersen, and Evert Vrugt (2017), "Carry," *Journal of Financial Economics*.

RavenPack (2019). "News Sentiment Analysis for Fixed Income Markets", *RavenPack Quantitative Research*, May 2019, **https://app.ravenpack.com/research/fixed-income-markets/**

Terms of Use

This White Paper is not intended for trading purposes. The White Paper is not appropriate for the purposes of making a decision to carry out a transaction or trade. Nor does it provide any form of advice (investment, tax, legal) amounting to investment advice, or make any recommendations regarding particular financial instruments, investments or products. RavenPack may discontinue or change the White Paper content at any time, without notice. RavenPack does not guarantee or warrant the accuracy, completeness or timeliness of the White Paper

You may not post any content from this White Paper to forums, websites, newsgroups, mail lists, electronic bulletin boards, or other services, without the prior written consent of RavenPack. To request consent for this and other matters, you may contact RavenPack at research@ravenpack.com.

THE WHITE PAPER IS PROVIDED "AS IS", WITHOUT ANY WARRANTIES. RAVENPACK AND ITS AFFILIATES, AGENTS AND LICENSORS CANNOT AND DO NOT WARRANT THE ACCURACY, COMPLETENESS, CURRENTNESS, TIMELINESS, NONINFRINGEMENT, TITLE, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE OF THE WHITE PAPER, AND RAVENPACK HEREBY DISCLAIMS ANY SUCH EXPRESS OR IMPLIED WARRANTIES. NEITHER RAVENPACK NOR ANY OF ITS AFFILIATES, AGENTS OR LICENSORS SHALL BE LIABLE TO YOU OR ANYONE ELSE FOR ANY LOSS OR INJURY, OTHER THAN DEATH OR PERSONAL INJURY RESULTING DIRECTLY FROM USE OF THE WHITE PAPER, CAUSED IN WHOLE OR PART BY ITS NEGLIGENCE OR CONTINGENCIES BEYOND ITS CONTROL IN PROCURING, COMPILING, INTERPRETING, REPORTING OR DELIVERING THE WHITE PAPER. IN NO EVENT WILL RAVENPACK, ITS AFFILIATES, AGENTS OR LICENSORS BE LIABLE TO YOU OR ANYONE ELSE FOR ANY DECISION MADE OR ACTION TAKEN BY YOU IN RELIANCE ON SUCH WHITE PAPER. RAVENPACK AND ITS AFFILIATES, AGENTS AND LICENSORS SHALL NOT BE LIABLE TO YOU OR ANYONE ELSE FOR ANY DAMAGES (INCLUDING, WITHOUT LIMITATION, CONSEQUENTIAL, SPECIAL, INCIDENTAL, INDIRECT, OR SIMILAR DAMAGES), OTHER THAN DIRECT DAMAGES, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. IN NO EVENT SHALL THE LIABILITY OF RAVENPACK, ITS AFFILIATES, AGENTS AND LICENSORS ARISING OUT OF ANY CLAIM RELATED TO THIS AGREEMENT EXCEED THE AGGREGATE AMOUNT PAID BY YOU FOR THE WHITE PAPER. JURISDICTIONS DO NOT ALLOW THE EXCLUSION OR LIMITATION OF LIABILITY FOR DAMAGES OR THE EXCLUSION OF CERTAIN TYPES OF WARRANTIES, PARTS OR ALL OF THE ABOVE LIMITATION MAY NOT APPLY TO YOU.

These Terms of Use, your rights and obligations, and all actions contemplated by these Terms of Use will be governed by the laws of New York, NY, USA and You and RavenPack agree to submit to the exclusive jurisdiction of the New York Courts. If any provision in these Terms of Use is invalid or unenforceable under applicable law, the remaining provisions will continue in full force and effect, and the invalid or unenforceable provision will be deemed superseded by a valid, enforceable provision that most closely matches the intent of the original provision.

RavenPack

RavenPack is the category leader in Alternative Data for the corporate and financial sectors since 2003. We develop AI technology to process content at scale, including one of the largest training sets for NLP applications. Our clients include the most successful hedge funds, banks, and asset managers in the world.

For more information about RavenPack, including trials, please contact us at:

Sales Team: ***sales@ravenpack.com***

www.ravenpack.com

Technical Support: ***support@ravenpack.com***

info@ravenpack.com

Americas: ***+1 646 277-7339***
EMEA: ***+44 (0) 20 3290 7395***
APAC: ***+65 3138 4809***