

The Journal of Portfolio Management

www.ijjpm.com

QUANTITATIVE STRATEGIES:

SPECIAL ISSUE
2017

FACTOR INVESTING

Contrarian Factor Timing
is Deceptively Difficult

CLIFFORD ASNESS, SWATI CHANDRA, ANTTI ILMANEN,
AND RONEN ISRAEL

Contrarian Factor Timing is Deceptively Difficult

CLIFFORD ASNESS, SWATI CHANDRA, ANTTI ILMANEN,
AND RONEN ISRAEL

CLIFFORD ASNESS
is managing principal of
AQR Capital Management
in Greenwich, CT.
cliff.asness@aqr.com

SWATI CHANDRA
is a vice president at AQR
Capital Management in
London, United Kingdom.
swati.chandra@aqr.com

ANTTI ILMANEN
is a principal at AQR
Capital Management in
London, United Kingdom.
antti.ilmanen@aqr.com

RONEN ISRAEL
is a principal at AQR
Capital Management in
Greenwich, CT.
ronen.israel@aqr.com

The growing popularity of factor investing has led to worries that factors may be overvalued today, thereby posing risks to investors in these strategies.¹ This begs the question—just how rich, if at all, are style premia factors?² And should investors hold off on investing in rich factors until they cheapen? We address these two questions in this article. First, we show value spreads for some well-known styles and find that while a few styles are expensive relative to history, as a group they are not, and none are at bubble-like (or the opposite) levels. Second, we examine tactical style timing strategies in a multi-style framework, asking whether dynamic allocations can improve the performance of a diversified multi-style portfolio. We compare the impact of value timing with that of strategic exposure to value itself. We find lackluster results here—strategic diversification turns out to be a tough benchmark to beat.

While we discuss only factors or style premia here, and use those terms interchangeably, our findings hold for smart beta, too. Smart beta is another form of factor investing, one that generally involves the simplest versions of known factors implemented in a long-only portfolio and, to date, mostly applied to individual stock selection. Asness [2016a] argued that these are essentially the same things.

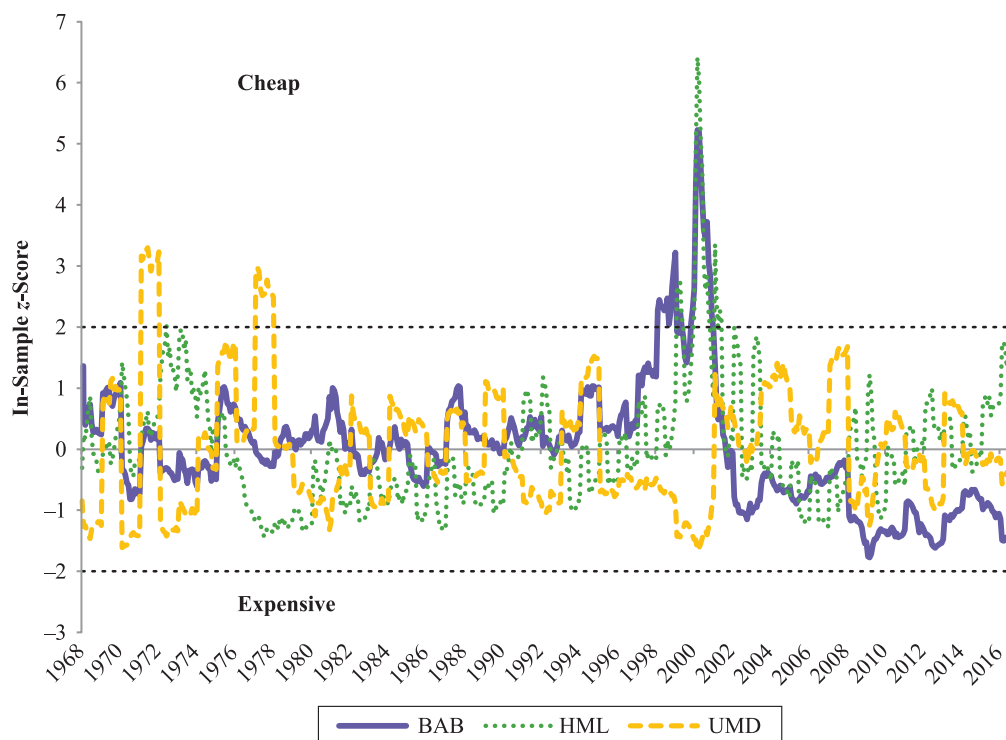
HOW RICH ARE FACTORS TODAY?

We start by introducing the concept of value spreads. Just as standard value measures quantify the cheapness of an asset relative to a fundamental anchor (for example, the CAPE of the market or the price-to-book ratio or price-to-sales ratio of a single stock),³ an equivalent metric of cheapness for a factor is the value spread, introduced by Asness et al. [2000], which is the ratio of some measure of value,⁴ say book to price, of the long side to the short side of the factor portfolio. When the ratio increases, it implies the factor is relatively cheaper and vice versa. As long as there is cross-sectional variance in a measure (e.g., firms have different book-to-price ratios if that's the value measure) you can always find stocks that are relatively cheap and expensive on that measure.⁵ The value spread measures “how much cheaper than normal” (or vice versa) are you finding such stocks today? Each value metric has a corresponding value spread. We will focus on some of the most common. To standardize the value spreads across time, we use z -scores, with a positive z -score indicating the factor is cheaper than its historical average and vice versa.

Exhibit 1 presents in-sample z -scores (henceforth referred to as STDs as they show how many standard deviation units each observation differs from the historical mean) of the book-to-price based value spreads for

EXHIBIT 1

Book-to-Price Spreads of U.S. Large-Capitalization Academic Styles, 1968–2016



Notes: The universe is U.S. large-cap stocks. The book-to-price ratio value spreads are standardized in-sample.

Source: Data from Xpressfeed, as of December 31, 2016.

three common, annually rebalanced academic factors applied to U.S. equities. High-minus-low (HML), up-minus-down (UMD), and betting-against-beta (BAB) represent the value, momentum, and low-beta (or defensive) styles, respectively.⁶ Throughout this article, we use the Asness and Frazzini [2013] specification of HML, but using the Fama–French specification would not alter any of the main results. We see that today, while, not surprisingly, some of these factors are cheaper and some richer compared with historical norms, none are near bubble-level extremes and collectively they do not paint a picture of very stretched valuations in either direction.⁷ In recent years, despite the proliferation of factor and “smart beta” investing, the spreads remain historically reasonable and exhibit a pattern of frequent mean-reversion, not steady richening in response to growing investor demand. Asness [2016a] found similar patterns and further showed how value spreads can vary widely with spread design decisions. For example, we find that using sales-to-price based

value spreads and industry-neutral variants of the factors, the BAB factor looks normal to cheap, whereas the value factor looks somewhat expensive, reversing direction from the findings using just book to price and not adjusting for industries.⁸ Book to price is a natural starting point given the literature, but including more metrics gives a broader picture and, in general, tends to make current factor valuations look more mundane.⁹ In particular, researchers who favor using composites of reasonable valuation measures would logically favor using similar composites to measure the richness or cheapness of factors. As of the end of our sample, such composites also generally yield much more mundane current readings than just book to price.¹⁰

Here we examine factors for large-capitalization U.S. stock selection. In Exhibit A1 and Exhibit A2 in Section A of the online appendix, we present multi-factor composite value spreads for other style premia specifications, namely, global large-cap stock-selection styles and global macro multi-asset styles going beyond

just equities, and we observe a similar lack of systematic richening over time and a similar “lack of anything particularly extreme” occurring at the time of this writing. That is, these findings of normalcy extend to stocks around the world and to other asset classes where styles like value, momentum, and defensive can also be applied.

TIMING EACH STYLE ON ITS OWN

We’ve established that value spreads for the main academic factors are not particularly extreme today. However, a related but separate question is whether taking account of and trading on these value spreads has been historically effective (i.e., if it were more extreme today should you vary your allocation to a style based on its valuation?). We now move on to this second question, focusing in particular on the incremental benefits of applying value-spread-based timing to a multi-style portfolio.¹¹ Please note that this is already a somewhat different question than asking whether timing has any predictive power at all for a single factor. The apparent mean-reverting nature of the value spreads in Exhibit 1 may suggest their potential use in tactical style timing, similar to how many use aggregate market valuations for broad market timing. But we know that contrarian market timing is very difficult,¹² and we find that successful contrarian timing of styles is at least as difficult.

Value timing in general relies on the premise that mean reversion in valuations is predictable and primarily due to changes in prices. However, we find that changes in style valuations are often driven by changes in portfolio positions or fundamentals, rather than changes in prices, especially for higher turnover styles like momentum or low beta. In contrast, the composition of the market portfolio changes very little through time. So, even if the value spread of a style mean reverts, the mean reversion could be for reasons other than price changes, thus not necessarily generating any gains from timing. For example, if the value spread of the value factor changes not because of price but because of book value changes, the factor investor didn’t necessarily make or lose money on this value spread change. Similarly, if the value spread of the momentum factor changes because of turnover in the stocks held long and short, mean reversion of the spread does not necessarily lead to any profit for a contrarian timer. Empirically, these mechanisms, unrelated to prices, do not destroy but do weaken the link

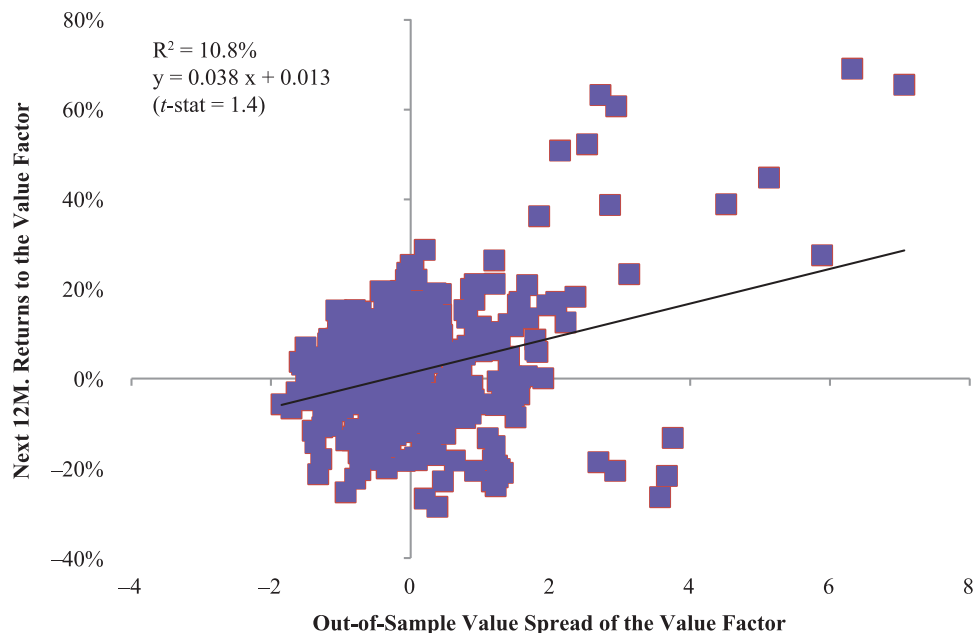
between spread mean reversion and profitable contrarian style timing.¹³ It would thus not be surprising to find the predictive power of value spreads for style returns to be even more limited than that of aggregate market valuations for market returns. In other words, while a well-constructed value factor (like the value spreads we use here) is likely to have *some* predictive power—value is after all a near universally effective strategy—the barriers to creating a very useful predictor are high.

Timing strategy backtests can be vulnerable to look-ahead biases. In particular, the use of in-sample spreads may overfit the past and underdeliver in the future. To avoid such biases in our analysis, at each point in time, we now use only the history available up to that point. We convert the value spreads to *out of sample* (OOS) *z*-scores using expanding windows with a minimum of 120 months of data.¹⁴ Exhibit 2 depicts the relationship between OOS value spreads of the annually rebalanced value factor and the next 12-month returns to value. The rather low R-squared of 0.10 (or correlation of 0.3) at the 12-month horizon indicates the expected positive but still tenuous relationship between OOS value spreads and future returns. Indeed, the statistically insignificant *t*-statistic of 1.4 confirms the frail nature of the relationship.¹⁵ Note again that we expect a positive relationship *ex ante*. There is near-unanimous support in the literature for the efficacy of value investing, and thus the value (or value spread) of anything should have some predictive power for that thing. The question is the strength of this relationship. Here we find that strength rather lacking.¹⁶ Although some may find this surprising, we find it at least consistent with other results. The market is simply a factor, market timing is notoriously difficult, and as we discussed earlier, the factors here are, if anything, less amenable to simple valuation-based timing than the more static market.

While Exhibit 2 gives a detailed visual example of the predictive relation for one style (value) and one horizon (next 12 months), Exhibit 3 succinctly shows the predictive correlations for three styles (value, momentum, and defensive, each annually rebalanced) across different return horizons. Positive correlations indicate a style is likely to have higher returns when cheaper, and vice versa. We observe lower correlations for momentum (UMD) than value (HML) and even persistent negative correlations for defensive (BAB).¹⁷ Furthermore, the correlations remain moderate at shorter horizons and may be overstated and misleading at longer horizons.¹⁸

EXHIBIT 2

Predictive Relation between Out-of-Sample Value Spreads and Next 12-Month Returns to Value, 1978–2016

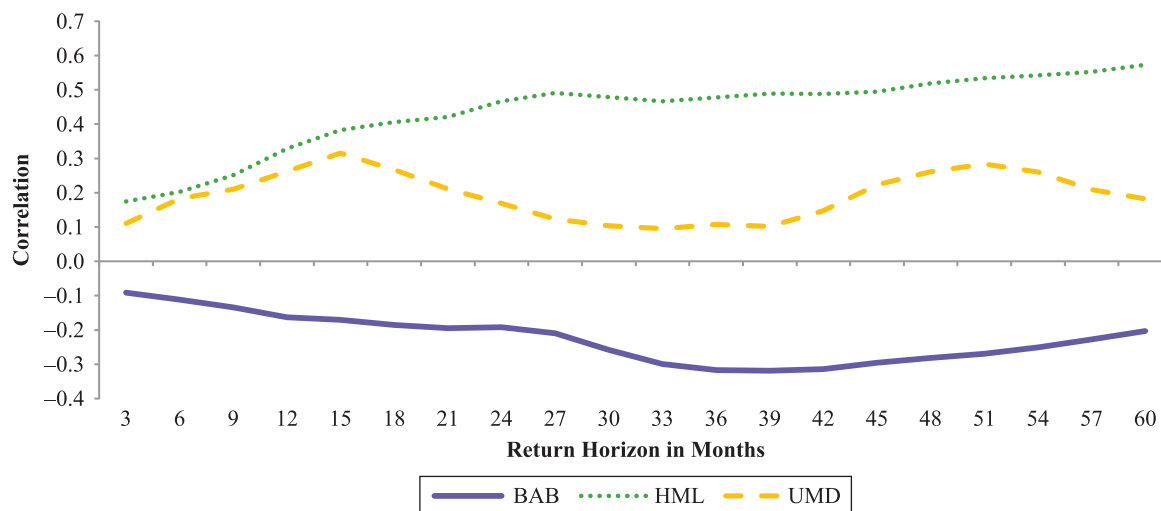


Notes: The universe is U.S. large-cap stocks. The value spreads are out-of-sample, standardized book-to-price ratio value spreads of the HML factor.

Source: Data from Xpressfeed, as of December 31, 2016.

EXHIBIT 3

Predictive Correlations of Out-of-Sample Value Spreads with Returns, 1978–2016



Notes: The universe is U.S. large-cap stocks. The value spreads are out-of-sample, standardized book-to-price ratio value spreads for the BAB, HML and UMD factors.

Source: Data from Xpressfeed, as of December 31, 2016.

For higher turnover factors in particular, the portfolio a few years out bears little resemblance to today's portfolio. So for some factors, it can be a spurious exercise to forecast returns five years out using the information we have today about today's portfolio. Asness [2016b] showed it can be nonsensical for momentum, rather silly for BAB, and even for value considerably dodgier than for the more stable aggregate stock market.¹⁹ As we shall see next, when implementing actual contrarian trading strategies based on value spreads, in the context of a multi-style portfolio, the mildly promising correlation picture (excluding BAB, which has a backward relationship) fails to translate into economically meaningful performance improvement.

TIMING IN A MULTI-STYLE PORTFOLIO

For someone invested in multiple styles, the relevant question is not whether any single style produces higher returns when cheaper, but whether value timing of multiple styles (also known as contrarian style rotation) reliably beats a strategic (i.e., not timed) multi-style allocation (in particular, one that includes the value factor itself). It is not surprising if a standalone style delivers higher future returns when cheaper, as value works on average and a cheap style has a value tailwind. When the value spread of a style indicates cheapness (richness), it is basically fighting value less (more) and is more (less) correlated with value at those times. Although we don't explore it here, this logic should extend to other styles too. Just as the valuation of a portfolio matters, so does its momentum or quality. It is not sufficient to ask if just the price is right, unless value is the only factor or style you believe in. A cheap style with a value tailwind may face a momentum or quality headwind and become even cheaper. Similarly, when momentum (or quality) is expensive, it follows that the correlation between momentum (or quality) with value is lower than usual or that the value factor is likely to have poorer than usual momentum or quality characteristics. As momentum and defensive styles work on average too, a style should work better when it has an above-average exposure to momentum or defensive. In other words, when any of the styles we believe in on average agree with each other (are more correlated than usual), one would expect better performance from them.

We digress here to comment on value and momentum in particular. If momentum is cheaper

than average, it implies cheap assets have outperformed recently (specifically, during the formation period of the momentum portfolio, which is 12 months for UMD) and are likely to load positively on momentum as well as value (or at least, these two factors will be less negatively correlated than usual). Thus, momentum loading on value is concomitant with value loading on momentum. So, if momentum does have higher returns when it is cheaper, it could be attributable to the momentum of the value factor or to the diversification benefits of combining value and momentum, and not just to the cheapness of the momentum factor. The story is similar for all factor pairs, though perhaps most acute for this one, as value and momentum are negatively correlated.²⁰

Simulation Methodology

To measure the incremental benefit from value timing, we compare the performance of a multi-style strategy with and without value timing each of the styles. Furthermore, to illustrate the contrast between the impact of value timing on strategies with and without pre-existing allocations to value, we simulate value timing on single-style strategies as well.

Our baseline "non-timed" case is an equal-weighted portfolio with a constant capital allocation to the style(s), which we refer to as the strategic weight(s).²¹ For single-style strategies (stand-alone value, momentum, or defensive), this weight is 100%, whereas for multi-style strategies that include two (or all three) styles, the strategic weights are 50% (or 33%) in each style. The portfolios are all rebalanced annually every January.

Our value timing signal is the OOS z -score of the book-to-price spread for each style. We vary the capital weight on each style between 50% and 150% of its strategic weight based on its OOS value spread.²² We do not short factors at any point in time, and we cap z -scores at ± 2 STD to prevent oversized bets. For multi-style strategies, to maintain a constant total dollar investment in the strategy, we prorate the tactical timing weights such that the tactical weights sum up to 100% across all styles at each point in time.²³ The contrarian timing strategy is rebalanced annually every January, as we expect value timing to be less effective at shorter holding periods due to the time typically required for mean reversion to occur.²⁴ As we shall see, we find results are largely robust to these choices and hold for

EXHIBIT 4

Value Timing Performance Summary, 1978–2016

	Non-Industry-Neutral				Industry-Neutral			
	Gross Returns		Gross Sharpe Ratios		Gross Returns		Gross Sharpe Ratios	
	Non-Timed	Value Timed	Non-Timed	Value Timed	Non-Timed	Value Timed	Non-Timed	Value Timed
Single Style Portfolios (V/M/D)								
HML	1.5%	2.2%	0.13	0.16	1.1%	1.4%	0.12	0.13
UMD	3.5%	4.0%	0.26	0.33	2.5%	2.7%	0.27	0.35
BAB	4.6%	3.9%	0.53	0.40	2.7%	2.7%	0.40	0.39
Style Portfolios with a Strategic Allocation to Value (V + M/D)								
UMD + HML	2.5%	2.9%	0.39	0.43	1.8%	1.9%	0.41	0.43
BAB + HML	4.1%	4.0%	0.52	0.49	2.5%	2.5%	0.46	0.45
Strategically Diversified Multi-Style Portfolios (V + M + D)								
HML + UMD + BAB	3.2%	3.4%	0.55	0.54	2.1%	2.3%	0.52	0.54

Notes: Returns are excess of cash and gross of transaction costs.

Source: Data from Xpressfeed, as of December 31, 2016.

other specifications of the value timing signal, holding period, and tilt.

We apply this predictive simulation to both industry-neutral and non-industry-neutral variants of value, momentum, and defensive (HML, UMD, and BAB, respectively).²⁵ The simulation runs from 1978 to 2016 to allow for the 120 months of history needed to create the OOS z -scores.

Exhibit 4 shows that value timing increases the return of the value (HML) portfolio but at the expense of increased volatility: its Sharpe ratio increases only slightly.²⁶ In the case of BAB, value timing decreases both return and Sharpe ratio. For the momentum (UMD) portfolio, we see an improvement in returns and Sharpe ratios, perhaps caused by the diversification benefits of adding a negatively correlated style, such as value, through the mechanism of value timing. However, if you wish to add value to a single factor portfolio like momentum, a strategic allocation to value may be more efficient than an implicit and intermittent exposure via value timing—the Sharpe ratio of value-timed momentum (0.33 for the non-industry-neutral case) is weaker than the Sharpe ratio of a constantly equal-weighted value and momentum strategy (0.39). Admittedly, applying value timing to the value and momentum strategy may still modestly improve upon the Sharpe ratio (to 0.43), but that remains lower than the Sharpe ratio (0.55) of the non-timed three-style portfolio that includes BAB.²⁷ Value timing this three-style portfolio

produces little to no improvement in either returns or Sharpe ratios. As the baseline portfolio gets more diversified, it becomes progressively harder for value timing to improve its performance.²⁸ Note these are *gross* Sharpe ratios and *gross* returns; all else equal, we would expect the higher turnover and transaction costs of tactical timing to result in a reduced *net* Sharpe ratio and *net* return advantage.

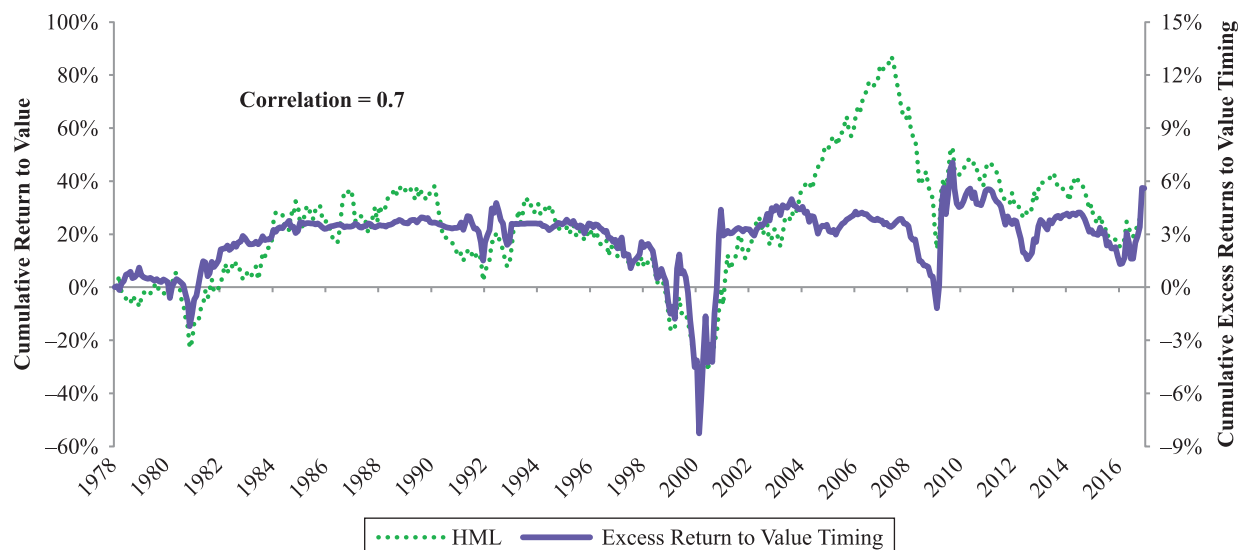
For simplicity, we have so far used a single value metric, book to price. In Section B of the online appendix, we run a similar simulation on global stock-selection style portfolios and multi-asset-class style portfolios using multi-factor composite value spreads that, as we have argued earlier, are better measures than any single measure standalone. We find even weaker results in that analysis, as we will discuss later in this article.

Tactical Value Timing versus a Strategic Allocation to Value

What explains the disappointing results, particularly in a multi-style context? As discussed previously, value timing entails an implicit bet on value itself as it involves a bigger bet on a style when it has greater exposure to the value factor. It is therefore not surprising that value timing is highly correlated to value. Specifically, the monthly return correlation between the excess returns of the value-timed portfolio (difference between the value-timed multi-style portfolio and

EXHIBIT 5

Co-Movement of Value and Value Timing, 1978–2016



Notes: The excess return to value timing is defined as the gross return of the value-timed multi-style portfolio minus the gross return of the non-timed multi-style portfolio that is equal-cap-weighted in the HML, UMD, and BAB factors.

Source: Data from Xpressfeed, as of December 31, 2016.

the non-timed multi-style portfolio) and returns to the signal used for value timing (HML in this case) is 0.7.²⁹ When using a multi-factor composite value spread as we do in Section B of the online appendix, the high correlation is to the multi-factor value composite.³⁰

Exhibit 5 plots the cumulative returns of value and of the excess returns of the value-timed multi-style portfolio versus the non-timed multi-style portfolio. We see that when value timing worked, as in the aftermath of the Tech Bubble bursting in 2000 and during the junk rally of 2009, so did the standard value factor. So, if a multi-style portfolio already includes value at optimally diversified levels, value timing the styles may increase value exposure to levels that undermine diversification, leading to weaker performance, particularly in a risk-adjusted sense. For many investors, the original intention of a multi-style allocation is to balance risk across multiple sources of return and capitalize on the power of diversification.³¹ Value timing a multi-style allocation may work against that very purpose by effectively increasing the allocation to value.

Any signal that has positive expected return and is uncorrelated to preexisting factors in the portfolio has the potential to boost performance if applied judiciously as a timing signal.³² Even high correlation to existing

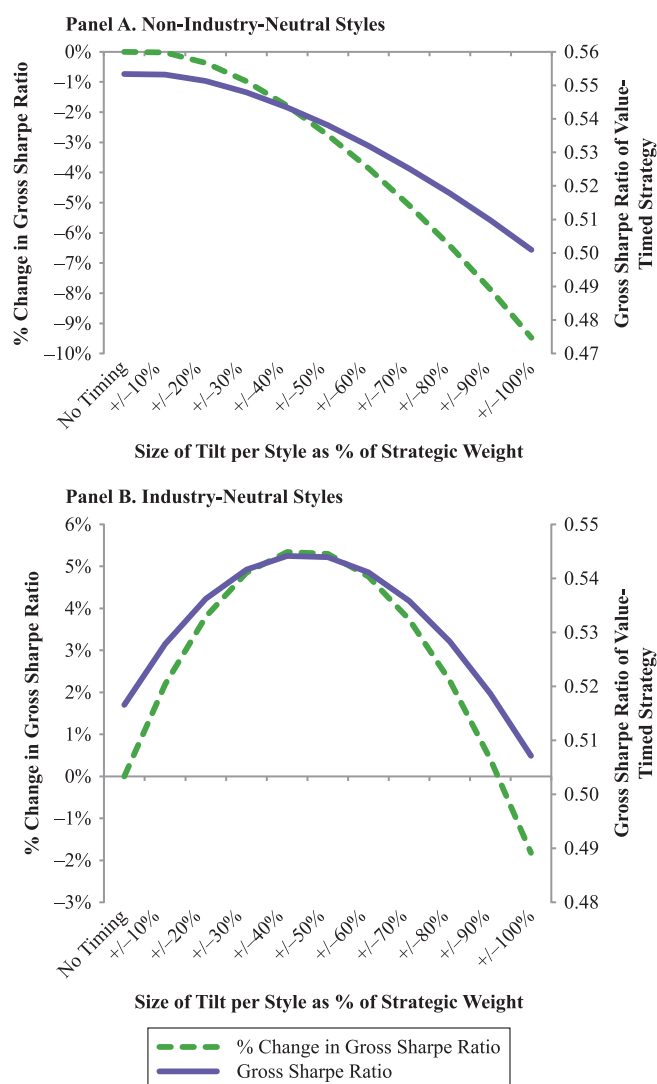
signals does not preclude a new timing signal from improving portfolio performance if used at a very moderate weight or if the power of the new timing signal is competitive with the existing signal itself. However, the greater the existing diversification in the portfolio and the greater the correlation of a tactical timing signal to the existing signals in the portfolio, the harder it is for tactical timing to improve risk-adjusted returns.³³ Portfolio math tells us that returns add linearly while risk adds quadratically. Hence, at larger tilts, the increase in risk from timing may be proportionately larger than any increase in return, resulting in lower risk-adjusted returns. Thus, intuitively, we expect smaller timing tilts to have more hope of boosting the Sharpe ratio than larger timing tilts (they are less likely to result in highly concentrated style bets that reduce diversification). This leads to the question: is there an optimal balance between tactical timing and diversification?

Value Timing with Different Timing Tilts

We attempt to empirically find that sweet spot in Exhibit 6, which plots the Sharpe ratio of a value-timed multi-style portfolio (that already includes value) as the size of the value timing tilt is varied. As before,

EXHIBIT 6

Gross Sharpe Ratio vs. Value Timing Tilt for Multi-Style Portfolios, 1978–2016



Notes: The initial portfolio is equal cap weighted in the HML, UMD, and BAB factors. Weights on the styles are varied in proportion to their OOS book-to-price spreads. The x-axis represents the size of the tactical tilt or maximum deviation from strategic weights allowed for any of the styles in the portfolio.

Source: Data from Xpressfeed, as of December 31, 2016.

styles are never shorted. For example, the $\pm 100\%$ tilt represents a timing strategy where a style is not held at all when unusually (more than 2 STD) rich. The $\pm 50\%$ tilt in Exhibit 6 matches the 50% to 150% maximum tilts applied in the simulation depicted in Exhibit 4. In line with our intuition, we see that in the case of

industry-neutral styles (Panel B in Exhibit 6), there is a minor improvement at the $\pm 40\%$ tilt, which is effectively a small tilt of 13.3% ($40\%/3$) for each style (due to the $1/3$ strategic allocation to each style). However, it is still negligible—gross Sharpe ratio increases from 0.52 to 0.54 and gross return, not shown, increases by 20 bps annually. If using the non-industry-neutral styles (Panel A in Exhibit 6), we see no benefit at even lower tilts.

This lackluster performance may stem from both the high pre-existing diversification³⁴ in the non-timed portfolio and the weak performance of HML itself, which, like many value factors, has had a bad draw in recent decades.³⁵ If the value factor has had difficulty stand alone, our first guess would be it would also be a difficult period using the corresponding value spread to time other factors. Value signals with higher Sharpe ratios will likely prove to be better timing signals (but also better competitors to timing as static strategic value factors). However, we find similarly weak results when using sales to price. All things considered, we find that if a value signal is really powerful, it may be most efficiently harnessed by introducing it into the portfolio via a constant strategic allocation rather than via an intermittent exposure through value timing of the factors themselves. Using that same signal for value timing may, essentially, be largely redundant (the timing induced value exposure is highly correlated with the strategic factor itself).

Value Timing at Longer Holding Periods

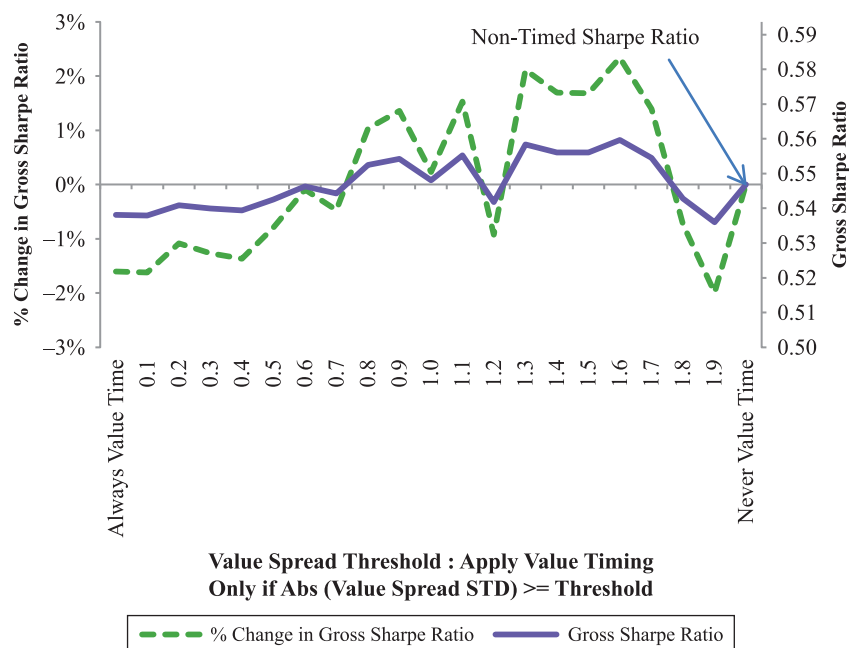
We now explore some other possibilities. First, contrarian style timing may demand greater patience—what if we use longer return horizons? Nope. Repeating the simulation with three-year holding periods leads to similarly weak results.³⁶ Similar to the annual rebalanced results, we find that value timing fails to noticeably increase either gross returns or Sharpe ratios for the multi-style strategy that includes HML, UMD, and BAB, with gross Sharpe ratios decreasing from 0.49 to 0.47 and gross returns remaining on par at 3.1%. For the single-style strategy of HML, it increases the return but not the Sharpe ratio, whereas for the single-style strategy of BAB, it decreases both the return and the Sharpe ratio.

Value Timing Only at Extreme Valuations

Next, what if value timing were applied only more selectively, say at extreme valuations? We intuitively

EXHIBIT 7

Value Timing Only at Extreme Valuations for a Multi-Style Portfolio, 1978–2016



Notes: The initial portfolio is an equal-cap-weighted portfolio of HML, UMD, and BAB. Weights on the styles are varied in proportion to their out-of-sample book-to-price spreads. The x-axis represents the minimum value spread threshold at which value-spread-based timing is applied.

Source: Data from Xpressfeed, as of December 31, 2016.

expect that minor fluctuations in value spreads might be noise and filtering these out should improve the value timing signal. To investigate this, we modify our simulation so as to apply value timing only when value spreads exceed (are either cheaper or richer than) a certain threshold. When the OOS value spread exceeds the threshold, the style weight is varied in proportion, as in the earlier simulation. If the OOS value spread is under the threshold, however, no value timing is applied and the style is held at its strategic weight. The higher the threshold, the more likely it is that valuations are actually stretched, but the fewer the instances where value timing is triggered and style weights varied from their strategic weights. The lower the threshold, the higher the number of instances where value timing is applied, but the greater the proportion of these instances that are likely to be noise.

Exhibit 7 shows the change in Sharpe ratios as the timing threshold is varied from 0 to 2 STD (because the spreads are capped at ± 2 STD, a threshold greater than 2 STD is equivalent to never using value timing).

As conjectured, we see the timed strategy Sharpe ratio improves as we increase the threshold, but it is a very modest improvement. The timed strategy Sharpe ratio barely exceeds the Sharpe ratio of the non-timed. In fact, increasing the threshold further leads to a slight drop in the Sharpe ratio. As we saw in Exhibit 5, when value timing works, value in general tends to work, and the multi-style portfolio's strategic allocation to value is sufficient to counteract exposure to overly expensive assets.

Value Timing Using Multi-Factor Value Composites and Multi-Asset-Class Styles

Finally as one last robustness check, in Exhibit 8, we explore contrarian style timing in other investable universes and asset classes including six global stock selection strategies (three industry-neutral stock selection and three industry selection) and six global macro strategies. Importantly, the style premia strategies here are mostly ex-ante beta neutral and target ex-ante constant volatility, and the multi-style strategies also target

EXHIBIT 8

Multi-Asset Evidence on Contrarian Style Timing, 1995–2016*

	Average Gross Returns		Average Gross Sharpe Ratios	
	Non-Timed	Value Timed	Non-Timed	Value Timed
Stock Selection	5.8%	5.8%	1.53	1.45
Industry Selection	3.2%	3.3%	0.69	0.69
Macro	3.8%	3.5%	0.84	0.75
Overall	4.1%	4.1%	0.98	0.91

*Data for Japan and emerging equities are from 2002 and for emerging currencies from 2003.

Notes: All data are monthly. Strategy returns are monthly returns, gross of fees and transaction costs and excess of cash.

Sources: Financial data and prices from IBES, Bloomberg, Datastream, Consensus Economics, Xpressfeed, MSCI Barra, and Penn World tables, as of December 31, 2016.

ex-ante equal risk allocation across the styles, making them more diversified than the previously discussed equal-cap-weighted multi-style strategy using academic factors. The overall simulation methodology is similar to that described earlier and depicted in Exhibit 4 but with a few differences, as described in Section B of the online appendix. Notably, we use multi-factor composite value spreads that are arguably more holistic indicators of valuation and hence, presumably, better timing signals. For example, for stock selection styles, we use a composite value spread that includes five factors (book-to-price ratio [BP], cash-flow-to-enterprise-value ratio [CP], earnings yield [EP], forward earnings yield [FEP] and sales-to-price ratio [SP]). We observe that the broad empirical evidence in Exhibit 8 confirms our earlier results in Exhibit 4. We see no meaningful increases in either Sharpe ratios or returns. In fact, we find value timing is even weaker here, judging by the drop in the average Sharpe ratio of the multi-style strategy (from 0.98 to 0.91) on applying contrarian factor timing.

COMPARISON WITH OTHER RESEARCH

Our repeated lack of success in value timing, across multiple samples and specifications, underscores the difficulty of implementing a value-timing strategy that can meaningfully improve upon a diversified multi-style strategy that includes value.³⁷ Yet, it continues to hold

appeal for many. For example, Arnott et al. [2016] and Arnott, Beck, and Kalesnik (hereafter, ABK; [2016a, 2016b]) explored very similar ground but have a different narrative on the current levels of style value spreads and their potential use in style timing. We note several points that may explain how they arrive at different conclusions.

First, their readings on current value spreads are more extreme than ours due to the use of percentiles instead of z -scores. Given the positive skew of value spreads seen in Exhibit 1, percentiles are more likely to look extreme than z -scores. We choose to use z -scores as they indicate relative magnitude while percentiles do not, something particularly important for very skewed statistics (e.g., if the CAPE is in the 90th percentile of history that is nearly the same as saying it's 90% as expensive as the peak of the Tech Bubble!). We nevertheless note some similarities—ABK [2016a] also reported multi-factor value spreads that are far less extreme than when using single metric book-to-price spreads (although we'd argue their tone seems to remain rooted in the initial single-factor book-to-price findings). For example, ABK found value spreads using only book-to-price indicate extreme richness (98th percentile) for the profitability factor and extreme cheapness (6th percentile) for the book-to-price factor as of March 2016. But their using a multi-factor value spread instead reflects normal valuations (40th percentile for profitability and 67th percentile for the book-to-price factor).

Next, Arnott et al. [2016] highlighted the correlation between *in-sample* value spreads and future five-year returns, to show the returns forecasting ability of value spreads. As we see in our Exhibit 3, a few styles display moderate predictive correlations (using *out-of-sample* value spreads). We do not find this surprising, as value works and we intuitively expect styles to have higher returns when cheaper. However, generally speaking, long horizon correlations or regressions as observed in Exhibit 3, or as documented in Arnott et al. [2016] and ABK [2017], which are based on persistent regressors like the *level* of value spreads, must be taken with a pinch of salt as they can be overstated as shown in Boudoukh, Richardson, and Whitelaw [2008] and in Boudoukh, Israel, and Richardson [2017]. Moreover, Asness [2016b] also showed that long horizons are deeply flawed for some factors—for example, momentum or BAB—where the valuation signal itself is much less autocorrelated than for, say, the market or value factors. Indeed, even for the

value factor, five years can be a stretch as the signal still has significant turnover at these horizons.

Furthermore, in forecasting long-term expected returns for styles based on their valuations, ABK [2017] ignored exposures to other styles. While we agree that—all else equal—a factor should have higher expected returns when cheaper, on similar principles, the returns for any style are also driven by its loadings on other styles besides value. We further note that ABK's use of a non-beta-neutral low-beta style greatly muddies results, as a dollar-neutral low beta factor will, by construction, have a persistent negative market beta. Unsurprisingly, it has far inferior performance (0.1 Sharpe ratio) versus the beta-neutral BAB factor (0.5 Sharpe ratio) we use to represent the low beta style.

Moving on to the correlation of value timing to the value factor itself, we consistently find very strong correlations and believe this is exceptionally intuitive. For instance, if a factor like momentum or low beta looks cheaper today than normal, our strong intuition would be that the standard value factor is itself more correlated to momentum or low beta (or less negatively correlated) than usual. This intuition is so strong and obvious we'd be shocked if we didn't find it empirically. Luckily we, repeatedly and in a very robust set of attempts, find the predicted and obvious correlation. Still, ABK [2016b] claimed otherwise, presenting some empirics. However, they used a multi-factor composite value spread as the timing signal but showed the loading of the excess return of the value-timed portfolio versus the non-timed portfolio on the HML Fama–French factor in a Fama–French four-factor regression. This mismatched method is imprecise and understates the correlation as it uses a different value measure (here a composite) as a timing signal from the one being used to construct the value factor. In particular, the value measures they're using to time factors are measured using up to date price while HML Fama–French uses lagged price, thus further muddling the story.³⁸ Measured apples to apples, using the same valuation measure to time factors and for the non-timed long–short value factor (with or without lags but matching on this choice), we always find very high correlations between the excess return of a value-timed portfolio versus its non-timed equivalent and the value factor itself.

Can value timing ever work? There is one instance, in ABK [2016b], where a value-based multi-style timing simulation slightly increases Sharpe ratio (from 0.52

to 0.66) and increases returns at the expense of additional portfolio volatility. We don't rule out the possibility of value timing working under certain scenarios. For *any* portfolio that includes a sufficient number of factors at suboptimally diversified weights, it is easier for *any* timing signal to improve returns and possibly Sharpe ratios.³⁹ In comparison, across our multiple iterations of timing simulations, using a range of out-of-sample value spreads, asset classes, and backtest specifications, we never find any meaningful increase in gross Sharpe ratio or gross return for the multi-style portfolio. Thus, we find that isolated combinations that lead to improving risk-adjusted returns from timing are not robust. This reinforces the conclusion that choosing factors well and constructing a well-diversified, risk-controlled portfolio (what we term *craftsmanship alpha*) can be a more reliable source of return than factor timing.

To summarize, Arnott et al. [2016] and ABK [2016a, 2016b, 2017] depicted a very different picture on the current levels of value spreads due to their focus on book-to-price spreads alone, despite the more mundane readings they found using other value metrics. Furthermore, their measure of the predictive power of value spreads appears overstated due to their use of long-horizon regressions that are not statistically sound. It is entirely plausible that they found a few instances where contrarian timing improves gross returns and gross Sharpe ratios of a multi-factor portfolio. As big believers in the value factor ourselves, we expect value timing to work in certain circumstances. However, our own slew of trading simulations using out-of-sample value spreads across different holding periods, asset classes, and value timing signals fails to produce economically meaningful improvement in either gross returns or gross Sharpe ratios, underscoring the difficulty of successfully implementing contrarian factor timing. Finally, ABK's [2016b] measure of correlation between value timing and value understates the true correlation as it does not compare apples to apples—the value-timing signal used is a multi-factor value composite that uses current prices, and as such, correlations should be measured against this composite, not the HML Fama–French factor that uses lagged prices.

CONCLUSION

We find that despite the increasing popularity of factor investing, the best-known styles are not overly

expensive today. Still, the question of what to do when/if they get expensive/cheap is an interesting one. At first glance, standalone styles may seem to perform better when cheaper. This is to be expected as value investing has been shown to be ubiquitously effective. But actually implementing a successful contrarian timing strategy is harder in practice and against relevant alternatives (well-diversified portfolios). From a multi-style perspective, it is hard for contrarian style timing to meaningfully improve upon simple strategic diversification.

Value timing of factors is highly correlated to the value factor itself. This is both intuitively quite obvious and empirically strongly confirmed. Thus, while value timing of a factor may boost the performance of a single-factor strategy, especially a negatively-correlated factor like momentum, it is of little added benefit to a diversified portfolio that already includes a strategic allocation to value as it may result in a larger bet on value than intended and weaken performance due to forgone diversification.

Asset valuations indeed matter—yet again, we are big fans of the value factor—but so do asset quality and momentum, which is why we prefer value in conjunction with other styles. In particular, combining value with momentum mitigates exposure to both expensive styles and out-of-favor falling knives. Indeed, tactical timing using a combination of value and momentum (or even the other known factors) and/or applied only at extremes (extremes that we do not observe as of this writing) may have potential, a topic we may return to in the future.⁴⁰ Until then, our research supports the approach of sticking to a diversified portfolio of uncorrelated factors that you believe in for the long term,⁴¹ instead of seeking to tactically time them.

ENDNOTES

We thank Jacob Boudoukh, Andrea Frazzini, Thomas Maloney, Tobias Moskowitz, Lasse Pedersen, Scott Richardson, Rodney Sullivan, and Ashwin Thapar for helpful comments and suggestions and William Fu and Lei Xie for data assistance. The views and opinions expressed are those of the authors and do not necessarily reflect the views of AQR Capital Management, its affiliates, or its employees; do not constitute an offer or solicitation of an offer, or any advice or recommendation, to purchase any securities or other financial instruments and may not be construed as such.

¹See Arnott et al. [2016], who foresee the reasonable probability of a smart beta crash as a consequence of the growing

popularity of factor-tilt strategies. We later present a detailed discussion of our results versus Arnott et al. [2016] and Arnott, Beck, and Kalesnik (hereafter, ABK [2016a, 2016b, 2017]).

²Groups of factors driven by a common theme (e.g. value or momentum) are often called “styles.” Ilmanen, Israel and Moskowitz [2012] and Asness et al. [2015b] illustrated the existence of style premia across several asset classes and how to capture them. Value refers to the tendency for relatively cheap assets to outperform relatively expensive ones. Momentum refers to the tendency for an asset’s recent relative performance to continue in the near future. Carry refers to the tendency for higher-yielding assets to provide higher returns than lower-yielding assets. Defensive refers to the tendency for lower-risk and higher-quality assets to generate higher risk-adjusted returns. Each style premia factor is a hypothetical long-short (L/S) style portfolio formed by taking long positions in the assets with the strongest style attributes and short positions in the assets with the weakest style attributes.

³When using price to sales, or sales to price, the price actually refers to enterprise value, not just equity market capitalization.

⁴We prefer to use the ratio, not the difference, in value measures between the long and short sides of the style portfolio, as, unlike the difference, the ratio is unaffected by aggregate market valuations. However, for macro style premia, value measures such as real bond yield are likely to have near-zero or negative values, causing sign flips or large fluctuations in ratio value spreads, making them harder to compare across time. So, for macro style premia, we use the difference in value measures whereas for stock selection style premia, we use the ratio as most value measures tend to be positive. As noted by Asness [2016a], whether, even for equities, there should be a separate role for the difference, interpreted more as a “carry” than “value” factor, is an open topic for future research.

⁵See Cohen, Polk and Vuolteenaho [2003] for a further discussion on value spreads.

⁶In creating these long-short factors we use the methodology of Fama and French [1993] but we exclude the effect of smaller-caps by using only a large-cap U.S. stock universe. Each factor is capitalization-weighted long the 1/3 best stocks and short the 1/3 worst stocks, and rebalanced annually every January. In this article, HML refers to book to price on the lines of the annual HML-Devil factor as described in Asness and Frazzini [2013]. HML-Devil uses more timely market prices, unlike the Fama-French version of the HML factor that uses stale prices as it lags both book value and price. UMD refers to 12-month price momentum excluding the most recent month. The BAB factor is a beta-neutral factor that is capitalization-weighted long the top 1/3 lowest-beta and short the 1/3 highest-beta stocks. It is leveraged on the long side proportionate to the ratio of shrunk betas between

the long low beta and short high beta portfolios to make it ex-ante beta neutral, similar to the methodology used in Frazzini and Pedersen [2014]. For the purpose of creating value spreads for BAB, we use the unlevered weights as described in Ilmanen, Nielsen, and Chandra [2015].

⁷As of December 31, 2016, we observe HML at 0.0 STD, BAB at -1.2 STD, and UMD at -0.1 STD. The corresponding value spreads for the industry-neutral versions of HML, BAB, and UMD are +0.4 STD, -1.2 STD, and -0.7 STD, respectively. For comparison, the value spread of HML, calculated in the same manner, peaked at +6.4 STD in March 2000 during the Tech Bubble. The positive skew in value spreads seen in Exhibit 1 means that using percentiles instead of z -scores, or using logarithms of the ratio of the value spreads instead of the plain ratio, would depict a more extreme picture today. We choose to use z -scores rather than percentiles when standardizing value spreads across time, because z -scores reflect relative magnitudes while percentiles do not.

⁸As of December 31, 2016 the sales-to-price based value spreads are at 0.0 STD for BAB, -1.3 STD for HML, and -0.3 STD for UMD, and at +1.2 STD, -1.2 STD, and +0.13 STD for the industry-neutral variants of BAB, HML, and UMD, respectively.

⁹ABK [2016a] too found vastly different readings when using multi-factor value spreads versus using book-to-price spreads.

¹⁰There is nothing inherent to a multi-factor composite value spread that would lead to persistently more mundane readings as it is constructed by combining z -scores of individual value spreads and then again standardizing the composite versus its own history—it is just that book to price happens to yield more extreme pricing at the end of our sample. Research has shown that, in general, using an average of multiple measures results in a better and more stable measure than using a single measure alone. See Asness et al. [2015a] and Arnott, Hsu, and Moore [2005].

¹¹We are talking about timing factors like momentum, defensive, and yes value itself, using their respective value spreads. We apologize for this somewhat industry standard overtasking of the word “value”! And, yes, we might actually use the phrase “the value spread of value”....

¹²See Asness, Ilmanen, and Maloney [2017].

¹³See Ilmanen, Nielsen, and Chandra [2015] and Asness [2016b] for a further discussion on why the contemporaneous relation between factor returns and valuation changes can be quite loose.

¹⁴Exhibit 1 uses *in-sample* (IS) z -scores of value spreads as it merely displays value spreads (and would correspond to the OOS version at the very end of the period). Elsewhere in the article, we use out-of-sample value spreads to discuss the predictive ability of value spreads. IS value spreads are not used to evaluate predictive power as they use the entire

history of data to standardize value spreads, including data not available at the time, leading to a look-ahead bias.

¹⁵The t -statistics are corrected using the standard Newey–West method to account for the use of overlapping data.

¹⁶Repeating this exercise of regressing next 12-month returns on OOS value spreads for BAB and momentum, we observe a similarly insignificant (and backwards signed) t -statistic of -0.6 for BAB and a marginally significant t -statistic of 2.0 for momentum (we’ll soon touch upon the momentum case in more detail). It is notable that most of the predictive relations are driven by the extreme data points of the Tech Bubble—omitting the calendar year 2000 would actually lead to an R-squared of 0 for value and a statistically insignificant relation for momentum as well. One interpretation is that without seeing Tech Bubble extremes, we find basically no relationship.

¹⁷Alternate specifications using industry-neutral BAB result in slightly stronger or more positive correlations. However, the correlations never get very high and, as we shall see, also do not prove to be effective timing signals.

¹⁸Long horizon correlations as observed in Exhibit 3, or as documented in Arnott et al. [2016] and ABK [2017], can be overstated. See Asness [2016b], Boudoukh, Richardson, and Whitelaw [2008], and Boudoukh, Israel, and Richardson [2017] for further details.

¹⁹Asness [2016b] found that Shiller CAPE is almost 0.80 correlated with its own three-year lag but the book-to-price spread of the book-to-price factor is only about 0.25 correlated, which challenges the relevance of longer-term forecasting.

²⁰The correlation between the returns to HML and UMD is -0.5 in the period January 1968 to December 2016, across both the non-industry-neutral and industry-neutral variants. Asness, Moskowitz, and Pedersen [2013] found value and momentum negatively correlated across eight diverse markets and asset classes.

²¹As the academic factors have different and time-varying volatilities and correlations, there is an implicit amount of style timing in the baseline equal-capital-weighted strategy. In Section B of the online appendix, we show the results of a similar simulation using multi-asset style premia targeting *ex-ante* constant volatility and where the baseline case targets equal risk allocation across styles, as opposed to equal capital allocation. In this case, where the baseline portfolio is more precisely risk balanced and diversified, we find even weaker results to contrarian style timing.

²²As we show later, we vary the maximum tilt allowed in the simulation to see if our findings are sensitive to the amount of tilt applied. However, we find similarly weak results throughout.

²³Thus we disallow “leveraging” the whole, for example, borrowing to take more risk when factors are on average attractive, or holding cash when factors are on average unattractive. This is purely a test of rotation strategies. Exploring overall timing is an interesting avenue for the future. Asness [2016b] examined similar timing strategies but that allow leveraging the overall strategy and finds desultory results but there is more to examine here.

²⁴Value in general works better at longer holding periods as the lag allows it to pick up some momentum exposure (or, at least to avoid some negative momentum exposure). Indeed, using annual or even three-year rebalancing slightly improves the performance of value timing relative to using monthly rebalancing as in Asness [2016b], but results remain weak nonetheless. If anything the even weaker results to monthly rebalanced timing strategies are a somewhat purer test of contrarian timing.

²⁵The regular academic factors HML, UMD, and BAB do not control for industry risk and can be dominated by unintended and time-varying industry exposure. Hence, we create industry-neutral variants for each style to capture a purer form of the style premium without any industry bets. Both sets of academic factors are not constructed to be conditionally beta neutral, except for the BAB factors, which are leveraged on the long side proportionate to the ratio of shrunk betas between the long low beta and short high beta portfolios to make them ex-ante beta neutral.

²⁶There is certainly a debate to be had about return versus Sharpe ratio. We don’t engage in the debate here but simply report both.

²⁷These may generally be lower Sharpe ratios than many readers are used to as we restrict ourselves to large-capitalization stocks in the U.S.

²⁸Our overall results echo those of Asness, Ilmanen, and Maloney [2017] on market timing. Correlation evidence analogous to Exhibits 2–3 (especially if done in sample and over long horizons) make contrarian market timing look promising, yet when an actual contrarian trading rule is applied, the performance improvement is weak. Here, the bar is raised further for contrarian style rotation when the relevant benchmark is a strategically diversified multi-style strategy. Thus, contrarian tilting across style premia is even harder to do successfully than contrarian market timing. This occurs both because the Sharpe ratio of the strategic baseline is higher (so even a marginally positive Sharpe ratio addition will have less of an impact) and because the value-add of the timing methodology itself, the value spread, is strongly positively correlated with one of the strategically held factors.

²⁹In the simulation in this article, prorating the weights in the multi-style strategy back up to 100% to have a constant dollar investment at each point in time somewhat decreases the amount of value timing and hence the correlation with

value, because the total strategy investment across all three styles will not be lowered if all styles are expensive and one can be overweight a rich style if the other styles are even richer. If, however, the dollar investment is allowed to vary as in Asness [2016b] such that it is never overweight a rich style or underweight a cheap style, the correlation is even higher at around 0.9. Essentially, value-spread timing always wants to be a value bet, and the fewer constraints you put on it, the higher the correlation between value spread timing and the regular value factor.

³⁰To get a precise measure of the correlation, it is important to measure correlation versus the *same* signal used for value timing. ABK [2016b] used a multi-factor composite value spread as the timing signal but showed the loading of the excess return of the value-timed multi-factor portfolio versus the non-timed multi-factor portfolio on the HML Fama–French factor in a Fama–French four-factor regression (which not only uses *only* book to price but is constructed at a different lag).

³¹Admittedly, diversification reduces risk and increases risk-adjusted returns, without always increasing returns. In practice, however, a higher Sharpe ratio may also result in higher returns relative to a higher risk portfolio as less risky portfolios are less likely to face drawdowns that warrant costly intervention measures that reduce the returns of higher risk portfolios.

³²Asness, Ilmanen, and Maloney [2017] showed that applying a combined value and trend timing signals to the aggregate equity market may improve risk-adjusted performance when limited to smaller tilts.

³³See AQR Capital Management [2014], “Challenges of Incorporating Tactical Views.” Overcoming the tactical timing induced loss in diversification would require signals with higher Sharpe ratios.

³⁴The average pairwise correlation between HML, UMD, and BAB is -0.1 , with no single correlation exceeding 0.4 .

³⁵Lest readers get too negative a picture of the value style’s performance in recent decades, we note the following mitigating circumstances: 1) Other value measures have fared better than book to price, 2) value, like many other styles, works better in small caps than in large caps, which we study, and 3) the HML Fama–French factor has outperformed the HML–Devil factor because the former embeds some momentum, due to its use of lagged market prices. This also explains why the HML–Devil factor’s performance improves when combined with momentum.

³⁶In this simulation, we apply value timing every three years in January, using a staggered three-year rebalance schedule where effectively a third of the portfolio is rebalanced every year by averaging the three-year rebalance schedules starting in 1968, 1969, and 1970 so as to avoid results

specific to one particular three-year rebalance schedule. Although the HML and BAB factor portfolios are rebalanced every three years, momentum (UMD) is rebalanced every year as we observe reversal, not momentum, at multi-year holding periods.

³⁷Value timing applied outside a multi-style factor investing framework, for example, to deep value opportunities in a different cross-asset-class opportunity set (covering a large number of pairwise trades) may present a more positive story, as discussed in Liew, Pedersen, and Thapar [2017].

³⁸Asness and Frazzini [2013] studied this issue in depth.

³⁹As an aside, it is always easier for timing to increase returns rather than Sharpe ratio, as timing introduces additional volatility through reducing diversification. We prefer to use risk-adjusted returns as our chosen measure of out-performance, in line with well-grounded portfolio theory and because while returns matter, so do risk and the related likelihood of drawdowns. We'd agree that an investor who cared about total return and not risk-adjusted return would be more amenable to value-based timing, but also more amenable to momentum, low beta, and other forms of timing.

⁴⁰One particularly interesting possible future extreme would be to see such rich value spreads (not just low but extremely low spreads versus history) that they seem to indicate that the long-run opportunity has been arbitrated away. That is not something we see today, not even close, but could occur at some point in the future (ironically this would likely be after significant further positive performance to the slower moving factors as for slow moving factors cheapness is highly related to prior medium-term performance). Thus it's not even clear what a factor investor today should root for! We guess it depends on their time horizon. For example, should investors prefer relatively "normal" factor returns (which, as always, entail risk of medium-term poor performance) ad infinitum? Or do they perhaps prefer better than usual returns for some foreseeable future as the long-term is indeed really arbitrated away (again, unlike what we are observing today)?

⁴¹We have, throughout this article, assumed belief in the efficacy of the value, momentum, and low beta factors. Which factors you believe in and why (or whether you believe in any) is, of course, a very important but very different topic, one addressed in Asness [2015].

REFERENCES

- AQR Capital Management. "Challenges of Incorporating Tactical Views." *Alternative Thinking*, Fourth Quarter 2014.
- Arnott, R., N. Beck, and V. Kalesnik. "To Win With Smart Beta Ask If the Price Is Right." Research Affiliates, 2016a.
- . "Timing Smart Beta Strategies? Of Course! Buy Low, Sell High!" Research Affiliates, 2016b.
- . "Forecasting Factor and Smart Beta Returns." Research Affiliates, 2017.
- Arnott, R., N. Beck, V. Kalesnik, and J. West. "How Can Smart Beta Go Horribly Wrong?" Research Affiliates, February 2016.
- Arnott, R., J. Hsu, and P. Moore. "Fundamental Indexation." *Financial Analysts Journal*, Vol. 61, No. 2 (2005), pp. 83-99.
- Asness, C. "How Can a Strategy Still Work If Everyone Knows About It." AQR Capital Management, 2015.
- . "The Siren Song of Factor Timing." *The Journal of Portfolio Management*, Special QES Issue, Vol. 42, No. 5 (2016a), pp. 1-6.
- . "My Factor Philippic." AQR Capital Management, 2016b.
- Asness, C., and A. Frazzini. "The Devil in HML's Details." *The Journal of Portfolio Management*, Vol. 39, No. 4 (2013), pp. 49-68.
- Asness, C., A. Frazzini, R. Israel, and T. Moskowitz. "Fact, Fiction, and Value Investing." *The Journal of Portfolio Management*, Vol. 42, No. 1 (2015a), pp. 34-52.
- Asness, C., J. Friedman, R. Krail, and J. Liew. "Style Timing: Value versus Growth." *The Journal of Portfolio Management*, Vol. 26, No. 3 (2000), pp. 50-60.
- Asness, C., A. Ilmanen, R. Israel, and T. Moskowitz. "Investing with Style." *The Journal of Investment Management*, Vol. 13, No. 1 (2015b), pp. 27-63.
- Asness, C., A. Ilmanen, and T. Maloney. "Practical Market Timing: Sin a Little." *The Journal of Investment Management*, forthcoming 2017.
- Asness, C., T. Moskowitz, and L. Pedersen. "Value and Momentum Everywhere." *The Journal of Finance*, Vol. 68, No. 3 (2013), pp. 929-985.
- Boudoukh, J., R. Israel, and M. Richardson. "A Note on Factor Timing." Working paper, AQR Capital Management, 2017.

Boudoukh, J., M. Richardson, and R. Whitelaw. "The Myth of Long-Horizon Predictability." *Review of Financial Studies*, Vol. 21, No. 4 (2008), pp. 1577-1605.

Cohen, R.B., C. Polk, and T. Vuolteenaho. "The Value Spread." *The Journal of Finance*, Vol. 58, No. 2 (2003), pp. 609-641.

Fama, E., and K. French. "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics*, Vol. 33, No. 1 (1993), pp. 3-56.

Frazzini, A., and L. Pedersen. "Betting against Beta." *Journal of Financial Economics*, Vol. 111, No. 1 (January 2014), pp. 1-25.

Ilmanen, A., R. Israel, and T. Moskowitz. "Investing with Style." White paper, AQR, 2012.

Ilmanen, A., L. Nielsen, and S. Chandra. "Are Defensive Stocks Expensive? A Closer Look at Value Spreads." White paper, AQR, 2015.

Liew, J., L. Pedersen, and A. Thapar. "Deep Value." Working Paper, AQR Capital Management, 2017.

To order reprints of this article, please contact Dewey Palmieri at dpalmieri@iijournals.com or 212-224-3675.

The views and opinions expressed herein are those of the authors and do not necessarily reflect the views of AQR Capital Management, LLC, its affiliates or its employees. This document does not represent valuation judgements, investment advice, or research with respect to any financial instrument, issuer, security, or sector that may be described or referenced herein and does not represent a formal or official view of AQR.

Contrarian Factor Timing is Deceptively Difficult

Online Appendix

Clifford Asness, Swati Chandra, Antti Ilmanen, and Ronen Israel

Clifford Asness is managing principal of AQR Capital Management in Greenwich, CT.
Cliff.Asness@aqr.com

Swati Chandra is a vice president at AQR Capital Management in London, U.K.
Swati.Chandra@aqr.com

Antti Ilmanen is a principal at AQR Capital Management in London, U.K.
Antti.Ilmanen@aqr.com

Ronen Israel is a principal at AQR Capital Management in Greenwich, CT.
Ronen.Israel@aqr.com

APPENDIX A | Current Value Spreads for Multi-Asset Styles

We present value spreads for style premia using different specifications and asset classes, as defined in Ilmanen, Israel and Moskowitz (2012). Exhibit A1 shows in-sample standardized value spreads for global stock selection styles, while Exhibit A2 shows in-sample standardized value spreads for global macro multi-asset styles. For neither group do we find current value spreads unusually rich.¹

APPENDIX B | Contrarian Style Timing Simulations on Global Multi-Asset Style Premia

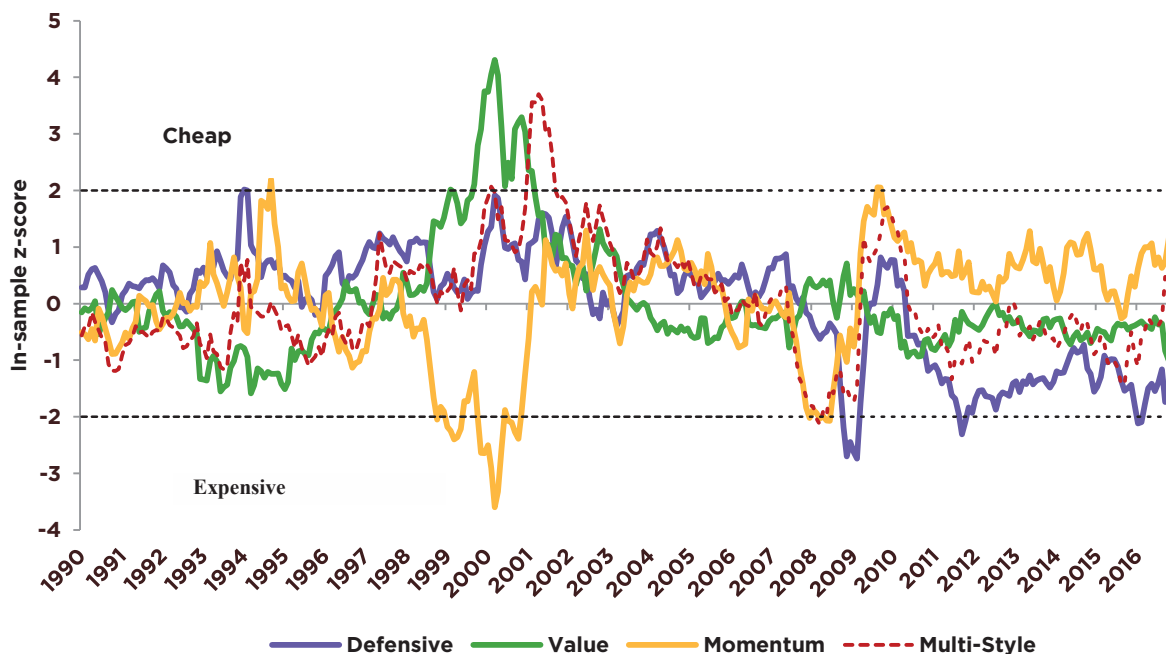
We examine contrarian style timing on a broader universe of multi-asset style premia that encompasses value, momentum, carry and defensive styles across six global stock selection (three industry-neutral stock selection, three industry selection) and six global macro strategies. (Here stock selection and industry selection are not applied only in the U.S. but also in Europe and Japan.) The stock selection and industry selection universe is again limited to large caps. Unlike the academic factors in the main text, the style premia here are multi-factor composites that are mostly ex-ante beta-neutral and target ex-ante constant volatility and ex-ante equal risk allocation across the styles in each strategy. We use a similar simulation

¹ As of December 31, 2016 value spreads are at -0.7 STD, +1.9 STD and -1.0 STD respectively for the global stock selection value, momentum and defensive styles in Exhibit A1. And +0.1 STD, -0.9, +0.9 and -0.7 STD respectively for the global macro value, momentum, carry and defensive styles in Exhibit A2.

methodology as described in the main text, adapted for multi-asset styles. The value timing signal is a multi-factor value composite using value measures relevant for each asset class. Due to a shorter history of data, the OOS value spreads are an expanding window z-score with a minimum window of 60 months. As the style portfolios and the value timing strategy are monthly rebalanced, and we expect value timing to work better at longer horizons, we use a 12-month average of the value timing signal.

Exhibit B1 compares the performance before and after value timing and shows results consistent with the findings in the main text. Value timing improves gross Sharpe ratios for 49% of single-style strategies (18 of 37 such strategies). However, as seen in Exhibit B1, these tend to be minor increases in both returns and Sharpe ratios, making the overall Sharpe ratio and return on average remain the same. Further, as these are *gross* returns and Sharpe ratio, they do not incorporate the higher turnover and transaction costs of tactical timing. More importantly, across asset classes, combining a style with value is noticeably better (0.68 Sharpe ratio) than value timing the style (0.63 Sharpe ratio). Once we apply value timing on strategies that already include value, value timing tends to detract from performance with the average Sharpe ratio decreasing from 0.98 to 0.91 as seen in the bottom table in Exhibit B1.

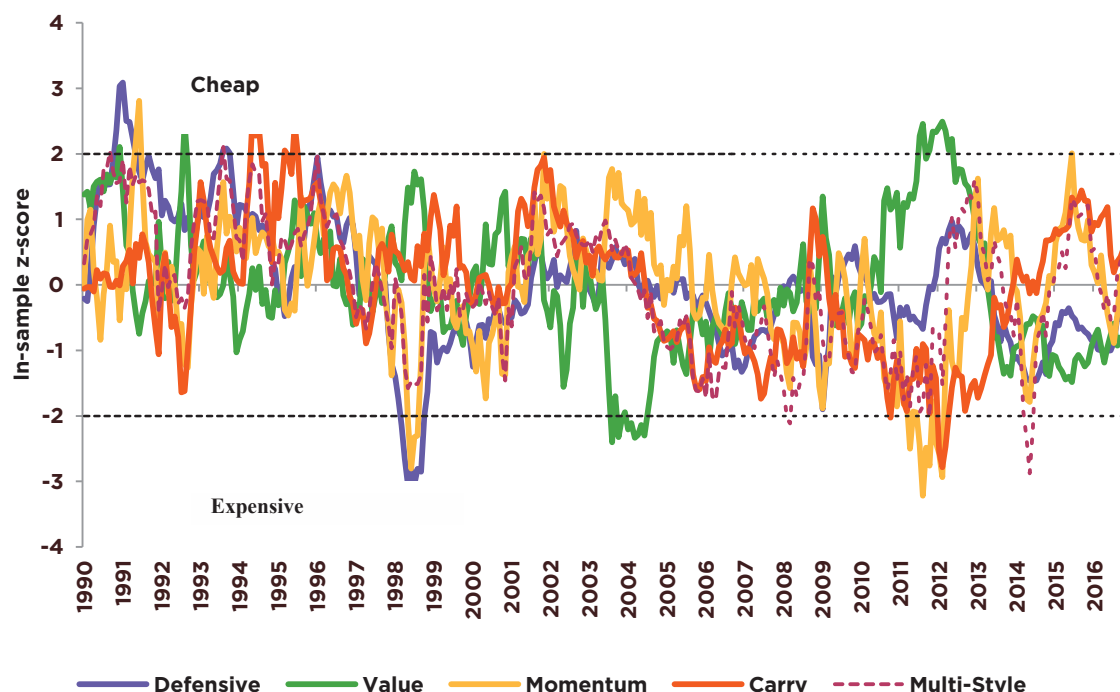
EXHIBIT A1 | Value Spreads for Hypothetical Global Stock Selection Styles, 1990*-2016



*Data for Japan from 1996.

Source: Data as of December 31, 2016. Financial data and prices are from Xpressfeed. Value spreads are a composite of ratio spreads using five value measures: book-to-price, cashflow-to-enterprise-value, earnings-to-price, forward-earnings-to-price and sales-to-price as described in Ilmanen, Nielsen and Chandra (2015). These style premia are captured by combining several indicators in each asset class and forming hypothetical long-short style portfolios that are rebalanced monthly while seeking to ensure the portfolio is market-neutral. The global stock selection universe comprises approximately 2,000 stocks across Europe, Japan, and the U.S.

EXHIBIT A2 | Value Spreads for Hypothetical Global Macro Multi-Asset Styles, 1990*-2016



*Data for Emerging Equities from 1996 and for Emerging Currencies from 1997.

Source: Data as of December 31, 2016. Financial data and prices from IBES, Bloomberg, Datastream, Consensus Economics, Xpressfeed, MSCI Barra and Penn World tables. All data is monthly.

Value spreads for the style premia are a composite of difference spreads using the respective value measures for each asset class as described in Ilmanen, Nielsen and Chandra (2015). These style premia are captured in numerous asset classes: stock selection, industry allocation, country allocation in equity, fixed income and currency markets, and commodities, by combining several indicators in each asset class and forming hypothetical long-short style portfolios that are rebalanced monthly while seeking to ensure the portfolio is market-neutral. The universes are as described: Developed Markets: Australia, Canada, Eurozone, Hong Kong, Japan, Sweden, Switzerland, U.K., U.S. Within Europe: Italy, France, Germany, Netherlands, Spain. Emerging Markets: Brazil, China, India, Israel, Malaysia, Mexico, Poland, Singapore, South Africa, South Korea, Taiwan, Thailand, Turkey. Bond Futures: Developed Markets: Australia, Canada, Germany, Japan, U.K., U.S. Emerging Markets: Czech Republic, Hong Kong, Hungary, Mexico, Poland, Singapore, South Africa, South Korea Yield Curve: Australia Germany, United States. Currencies: Developed Markets: Australia, Canada, Euro, Japan, New Zealand, Norway, Sweden, Switzerland, U.K., U.S. Emerging Markets: Brazil, Hungary, India, Israel, Mexico, Poland, Singapore, South Africa, South Korea, Taiwan, Turkey. Commodity Selection: Silver, copper, gold, crude, Brent oil, natural gas, corn, soybeans.

EXHIBIT B1 | Multi-Asset Evidence on Contrarian Style Timing, 1995*–2016

Single-Style Portfolios (V/M/C/D)

	Average Gross Returns			Average Gross Sharpe Ratios		
	Non-Timed	Value-Timed	Hit Rate	Non-Timed	Value-Timed	Hit Rate
Stock Selection	5.8%	5.9%	5 / 9	0.69	0.71	5 / 9
Industry Selection	3.2%	3.2%	6 / 9	0.35	0.37	7 / 9
Macro	3.6%	3.4%	8 / 19	0.36	0.36	6 / 19
Overall	4.0%	4.0%	19 / 37	0.44	0.44	18 / 37



Style Portfolios with a Strategic Allocation to Value (V + M/C/D)

	Average Gross Returns			Average Gross Sharpe Ratios		
	Non-Timed	Value-Timed	Hit Rate	Non-Timed	Value-Timed	Hit Rate
Stock Selection	6.2%	6.2%	3 / 6	1.29	1.23	2 / 6
Industry Selection	2.3%	2.3%	3 / 6	0.43	0.42	3 / 6
Macro	3.0%	2.8%	5 / 13	0.51	0.46	4 / 13
Overall	3.6%	3.5%	11 / 25	0.68	0.63	9 / 25



Strategically Diversified Multi-Style Portfolios (V + M + C + D)

	Average Gross Returns			Average Gross Sharpe Ratios		
	Non-Timed	Value-Timed	Hit Rate	Non-Timed	Value-Timed	Hit Rate
Stock Selection	5.8%	5.8%	2 / 3	1.53	1.45	0 / 3
Industry Selection	3.2%	3.3%	2 / 3	0.69	0.69	1 / 3
Macro	3.8%	3.5%	2 / 6	0.84	0.75	1 / 6
Overall	4.1%	4.1%	6 / 12	0.98	0.91	2 / 12

*Data for Japan and Emerging Equities from 2002 and for Emerging Currencies from 2003.

Source: Data as of December 31, 2016. Financial data and prices from IBES, Bloomberg, Datastream, Consensus Economics, Xpressfeed, MSCI Barra and Penn World tables. All data is monthly. Strategy returns are monthly returns, gross of fees and transaction costs and excess of cash. Hit Rate refers to the number of strategies where value timing increases the gross Sharpe ratio or return, divided by the number of strategies considered. Styles as defined in Exhibits A1 and A2 in Section A of the online appendix. Value spreads are a composite of ratio spreads for stock selection and industry selection strategies and difference spreads for global macro strategies using the respective value measures for each asset class as described in Ilmanen, Nielsen and Chandra (2015).